Phonetic effects on the timing of gestural coordination in Modern Greek consonant clusters

by

Jonathan Chung-Kay Yip

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in the University of Michigan
2013

Doctoral Committee:

Professor Patrice Speeter Beddor, Chair
Professor Julie E. Boland
Associate Professor Andries W. Coetzee
Professor San Duanmu

For my parents,

Lawrence and *Cecilia*,

and my brother,

Brian.

ACKNOWLEDGEMENTS

Foremost, I wish to thank my outstanding advisor and mentor, Pam Beddor. Countless times have you listened to my incessant doubts and told me precisely the right things to ground those concerns in reality. The conversations during our frequent, one-on-one meetings not only improved my theoretical understanding and scholarly writing, but also helped me manage the stress associated with teaching, career development, and family life. Your wisdom and patience are unparalleled, even during those grueling Thursday late-afternoon discussions! If I am ever aware of my strengths as a scholar, collaborator, technical expert, or educator, I sincerely owe that awareness all to you.

Thank you, too, to my terrific committee members, Andries Coetzee, San Duanmu, and Julie Boland, for your extremely useful input over the course of development of this dissertation. Each of you has given me access to new theoretical discussions, broadening my understanding of this work. I am also fortunate to have been supported other faculty and staff here at the University of Michigan—especially Sandy Petee, Nancy Herlocher, Ioulia Kovelman, Sam Epstein, Jeff Heath, Sally Thomason, Acrisio Pires, and Marlyse Baptista—who always made me feel at home on the fourth floor of Lorch Hall.

Also, thank you to those who collaborated with me on this dissertation, particularly my dedicated assistants, Tony Natoci and Amy Hemmeter, my language consultants, Evanthia Diakoumakou and Lavrentia Karamaniola, and talented linguists Alan Wrench, Ann-Michelle Tessier, Khalil Iskarous, and Jelena Krivokapić. This dissertation would not be possible without you. And thank you to the regular attendees to the Phonetics-Phonology Discussion (Phondi) group, who allowed me to regale them with the finer, complicated details of my experimental designs.

I would also like to thank my fellow graduate students, including Susan Lin, Anthony Brasher, Terrence Szymanksi, Kevin McGowan, Lauren Squires, Erica Beck, Stephen Tyndall, Tim Chou, Harim Kwon, Tridha Chatterjee, Michael Opper, Cameron Rule, Sujeewa Hettiarachchi, and Batia Snir. You have made our Department a brilliant and vibrant one! Special thanks to the members of my graduate cohort, David Medeiros, Eric Brown, and Joseph Tyler, and honorary cohort member, Autumn Fabricant, for the good humor and friendship we shared with each other and with others. I never felt alone knowing you were there.

To my friends at the Ella Jo Baker Graduate Cooperative, Denise Bailey, Katie Dover-Taylor, Josh Steverman, Gayle Myerson, Erin Van Campen, Julia Raskin, Mari Corella, Zach Singer-Leavitt, Paul Wentzell, Brian Devree, Nick Whitaker, John Bailey, Judith Schmidt, Mandy Kain, Natalie Matuszczak, Meredith Garry, and Abhinav Jain, thank you for experiencing the easy and difficult times with me. I have learned so much from each of you. We built one remarkable home together.

Lastly, to my family: through your unconditional love, I now understand the pride that you take in my accomplishments. My father, Larry Yip, has always encouraged me to follow my dreams for as long as I can remember. Thank you for setting me on the path to lasting fulfillment. My brother, Brian Yip, has been supportive, emotionally and financially, in life's most critical moments, rare as they are. By your example, I am able to stand on my own two feet. My mother, Cecilia Yip, never ceases to breathe meaning into my life. In your memory, I take no experience for granted and strive to be a better person each day. As your son, I feel blessed.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Figure

LIST OF APPENDICES

Appendix

CHAPTER I
Introduction

The central goal of this dissertation is to unite the various observations about gestural timing and coordination during the production of two-consonant clusters (CC) available in the current literature into a single project that addresses whether or not these articulatory, acoustic, and perceptual observations on intergestural coordination are motivated by the perceptual requirements of the listener or by inherent physical or mechanical characteristics of the human vocal tract, or both. This question is important to general theories that aim to model how a wide array of linguistic, paralinguistic, and biological information is integrated into the rich acoustic signal that results from speech production. If language structures produced by speakers are to be understood by listeners, then, arguably, the forms that speakers produced should be, at some level of representation, isomorphic with the forms on which listeners rely for perception (Goldstein & Fowler, 2003; Fowler & Galantucci, 2005).

A key issue for theories of speech production is the identification of the space in which phonetic productions are planned, and there is disagreement as to the nature of this planning space among theorists. Gesturalist theorists, such as proponents of Articulatory Phonology (Browman & Goldstein, 1992; Goldstein & Fowler, 2003; Goldstein, Byrd, & Saltzman, 2006), posit that phonetic planning is motoric in nature, and that listeners perceive the acoustic signal along gestural lines. Under such theories, gestures act as both the units of phonetic production and the objects of speech perception, and successful communication depends on whether speakers provide sufficient information about the gestures themselves in their phonetic productions (Löfqvist, 2010).

In other theories, such as the DIVA model (Directions Into Velocities of Articulators) (Guenther, 1995; Guenther *et al.*, 1998, 2006), the nature of the phonetic planning space is primarily auditory in nature, and the objects of production and perception are the acoustic targets shared among language users. Under such accounts, what counts as phonetic knowledge depends on the target acoustic profiles of speech sounds, rather than their articulatory features, and equivalence between the acoustic output of speakers and the acoustic representation of listeners is tuned via auditory feedback. In such a theory of speech production, speakers plan their articulatory patterns based on their knowledge of which patterns result in the best acoustics for their listener.

This dissertation contributes to the discussion on the nature of the phonetic planning space by attempting to tease apart the predictions made by gestural/biomechanical and acoustic/perceptual approaches to the study of speech production. In this study, I investigate the coordination of articulatory gestures in Modern Greek consonant sequences—C1-C2—that vary by four linguistic factors (within-word position, order of C1-C2 place of articulation, C1 manner, and C2 manner), whose influences on gestural coordination result in different sets of predictions within biomechanical and perceptual accounts.

1.1 Vocal-Tract Mechanics and Perceptual Recoverability

Biomechanical explanations for gestural coordination generally posit that variations in intergestural timing result from differences in the intrinsic motor characteristics of the speech articulators (Hardcastle & Roach, 1979; Chitoran & Goldstein, 2006). Because speech articulators such as the lips, the tongue tip, and the tongue dorsum may differ in the inherent speed at which they move and the extent to which they influence the movement of other articulators, timing patterns for gestural coordination will depend on the specific gestures involved. For example, the coordination of a bilabial plosive with a following alveolar plosive, as in [pt], may involve substantial temporal overlap between gestures because the lips and the tongue tip can move into their respective closed positions without restricting the other's gestural movement. However, the sequence of a velar plosive followed by a coronal plosive, as in [kt], might be expected to be produced with gestures that are less overlapped, given that tongue-dorsum raising toward the velum followed by apical raising toward the alveolar ridge requires two different constriction locations for the tongue. Consistent with this characterization, substantially more overlap between C1 and C2 has been observed in labial-coronal sequences than in dorsal-coronal sequences (Kühnert *et al.*, 2006).

However, in recent years, researchers have also found support for the notion that the timing patterns of phonetic gestures are influenced by the perceptual recoverability of the intended sequence. Specifically, gestures seem to be timed and coordinated so as to achieve sufficient, or perhaps optimal, information in the acoustic signal that serves as input to the listener (Silverman, 1995; Wright, 1996; Chitoran *et al.*, 2002; Kochetov, 2006). If gestural overlap might yield poor perceptibility, then articulatory planning must (arguably) compensate, for example, by reducing gestural overlap, so as to ensure successful perception. Thus, for sequences in which gestural overlap could acoustically mask the identity of the target sounds, speakers appear to be more likely to minimize overlap (Chitoran *et al.*, 2002; supporting acoustic data in Wright, 1996 and Chitoran, 1999). Such accounts generally predict that specific articulatory timing patterns may be driven in part by speakers' need to communicate effectively to their listeners.

1.2 Gestural Coordination in C1-C2 Sequences

C1-C2 sequences (henceforth "CC sequences") are of special interest to the study of gestural coordination during speech production. Languages tend to restrict the types of CC sequences that can occupy the initial position of words (or syllables), and speakers of such languages have difficulty perceiving productions of phonotactically illegal CC sequences as sequences of two consonants, instead often hearing two consonants separated by an epenthetic vowel (English: Moreton, 2002, Davidson, 2006; French: Hallé *et al.*, 1998; Japanese: Dupoux *et al.*, 1999, 2001). The inventory of possible CC sequences in a language may also be limited by the place combinations that may occur (e.g., dorsal-labial sequences such as [kp] and [gb] are unattested in Modern Greek, but dorsal-coronal sequences such as [kt] and [ɣð] occur), or the combinations of manner of articulation that are possible (e.g., English and German do not allow plosive-plosive sequences word-initially). If the restrictions on phonotactically-allowed CC sequences across the world's languages arise from the articulatory and/or perceptual pressures that limit the use of dispreferred CC sequences, then the productions of CC sequences in languages allowing for large sets of CC sequences should provide evidence of such perceptual or articulatory influences.

In the body of literature on gestural coordination in CC sequences, four phonetic factors—position within word, order of C1-C2 place of articulation, C1 manner, and C2 manner—are among those that have been found to affect the coordination of gestures in CC sequences during production. These effects, whose explanations in terms of biomechanics and/or perceptual recoverability are discussed in Sections §1.3-1.6, are as follows:

1. *Position in word*: The position of a CC sequence within a word influences the degree of gestural overlap between C1 and C2; namely, there is less overlap in word-initial contexts than in word-medial contexts (Wright, 1996; Chitoran, 1999; Chitoran *et al.*, 2002).
2. *Place order*: For both initial and medial positions, the order of place of articulation in a CC sequence influences the degree of gestural overlap between C1 and C2; namely, there is less overlap in back-to-front sequences than in front-to-back sequences (Byrd, 1992, 1996; Zsiga, 1994; Peng, 1996; Wright, 1996; Surprenant & Goldstein, 1998; Chitoran, 1999; Chitoran *et al.*, 2002; Chitoran & Goldstein, 2006).
3. *C1 manner*: The manner (fricative or plosive) of C1 in a CC sequence has been found to influence the degree of gestural overlap between C1 and C2, that is, there is less overlap in clusters containing a plosive C1 than in clusters containing a fricative C1 (Byrd, 1996; Kühnert *et al.*, 2006).
4. *C2 manner*: The manner (plosive or liquid) of C2 in a CC sequence influences the degree of gestural overlap between C1 and C2; namely, there is less overlap in clusters containing a plosive C2 than in clusters containing a liquid C2 (Chitoran & Goldstein, 2006, as presented in Chitoran & Cohn, 2009; Kühnert *et al.*, 2006).

While these four effects are not the only intergestural-timing effects that have been reported for CC sequences (e.g., syllable affiliation affects word-medial CC sequences: Byrd, 1996; Kochetov, 2006), they are the main intergestural-timing effects for which explanations both in terms of articulatory constraints and perceptual recoverability requirements have been considered. In the sections that follow, I discuss the acoustic, articulatory, and/or perceptual evidence for each of these effects. Biomechanical explanations for these effects generally focus on how the motor and/or aerodynamic characteristics of the movement of the speech organs influence intergestural motor coordination during production. Perceptual-recoverability accounts, on the other hand, generally interpret such timing effects as the product of speakers' awareness of their listeners' perceptual needs; that is, speakers adjust the timing of their productions so as to make phonetic information better available to listeners and less susceptible to acoustic masking effects.

1.3 Effect of Within-Word Position on CC Overlap

Prior studies (Wright, 1996; Chitoran, 1999; Chitoran *et al.*, 2002) have found that word-initial CC sequences tend to be produced with less overlap (i.e., more lag) between C1 and C2 than word-medial CC sequences. This effect has been interpreted as primarily a result of differing perceptual limitations on CC-sequence production in the two positional contexts. For word-medial sequences, listeners have access to both transitional cues into C1 and transitional cues out of C2, while for word-initial sequences, the transitional cues into C1 are absent. If acoustic cues for C1 constriction release were unavailable due to gestural overlap with C2 during production, C1 would be rendered with fewer cues to its identity in word-initial than in word-medial contexts. Such an interpretation of the position effect, however, would seem to be mitigated by the likelihood that a final vowel (or, in some cases, a final consonant) in a word immediately preceding the word-initial CC sequence may also carry transitional cues into C1, in which case recoverability of C1 might not be markedly different in initial and medial sequences.

Evidence of the within-word effect comes from Wright (1996), who acoustically analyzed word-initial and word-medial plosive-plosive sequences [pt pk tp tk pɗ tɓ kɗ] produced by five native speakers of Tsou. He found a higher rate of C1 release in word-initial environments (100%) than in word-medial environments (58%). Wright interpreted this lack of measurable C1 release in word-medial position as a result of insufficient pressure buildup behind the C1 constriction due to increased temporal overlap between C1 and C2 gestures word-medially.

Similarly, Chitoran (1999) acoustically measured release-burst and inter-burst intervals (IBIs) between C1 and C2 in plosive-plosive sequences [dg gd t^hk^h k^ht^h] produced by two native speakers of Georgian. She found that, in cases in which both C1 and C2 release bursts were present, the IBI was approximately 30% longer word-initially (mean: 126 ms) than word-medially (mean: 97 ms). These durational measurements are suggestive of greater articulatory overlap in word-medial plosive-plosive sequences than in corresponding word-initial sequences because shorter C1-C2 burst IBIs indicate longer intervals during which C1 and C2 gestural movements and plateaus co-occur.

In a subsequent study by Chitoran *et al.* (2002) that used electromagnetic midsagittal articulography (EMMA), Georgian plosive-plosive clusters [p^ht^h bg t^hb dg gb gd] were found to have greater C1-C2 gestural overlap word-medially than word-initially for one speaker, but not the other speaker in that study. Chitoran *et al.* note work on spoken word recognition (Marslen-Wilson, 1987) supporting the view that the onsets of words are important to lexical access and that minimizing gestural overlap at word onset rather than in the middle of the word will more effectively aid perceptual recoverability by the listener.

1.4 Effect of Place Order on CC Overlap

For the effect known as "place order" on gestural coordination in CC sequences, the anterior-posterior location of the C1 constriction in relation to that of the C2 constriction has been shown to influence both gestural timing and perceptual recoverability. One articulatory finding is that front-to-back plosive-plosive sequences exhibit greater overlap than do their back-to-front counterparts. In studies that pursue a perceptual-recoverability explanation for this effect (Wright, 1996; Chitoran *et al.*, 2002), this outcome is claimed to result from differences in the recoverability of acoustic information for the two plosives when the gestures for both sounds overlap temporally. Under such an account, speakers are more likely to reduce the overlap between gestures in back-to-front than front-to-back sequences, because the release of a front constriction during a back closure in front-to-back sequences should be audible, whereas a front constriction would mask a back release during back-to-front sequences.

An important point for the current study is that perceptual recoverability does not predict the same effect of place order for CC sequences that are not a combination of two plosives. For example, in a plosive-liquid sequence, such as [kl], sufficient cues to C1 place and manner should always be available, even with extensive intergestural overlap between C1 and C2. Because acoustic cues for the plosive should be present in the formant transitions into the following liquid, the perceptibility of the plosive is expected to change relatively little as overlap between the two gestures increases. Thus, there should be relatively little perceptual motivation for a back-to-front sequence to be less overlapped than a front-to-back sequence when C1 and/or C2 are not plosives.

Some studies on the effect of place order, i.e., Chitoran *et al.* (2002) and Kühnert *et al.* (2006), have focused on gestural coordination in CC sequences in which C2 is held constant, namely, as a coronal. Thus, for these investigations, "front-to-back" is a labial-coronal combination (e.g., [pt pl pɾ]), and "back-to-front" is a dorsal-coronal combination (e.g., [kt kl kɾ]). These researchers' proposed account for less overlap in dorsal-coronal than labial-coronal CC sequences has more to do with which articulators are involved than the general place order of constrictions: the former involve the coordination of two separate articulators, the lips and the tongue tip, whereas the latter involve constrictions made by posterior and anterior portions of the same articulator (the tongue). When "place order" differences involve contrasts in the degree of independence of the articulators involved, front-to-back coronal-dorsal sequences (e.g., [tk dɡ]), like their back-to-front dorsal-coronal counterparts ([kt ɡd]), should involve relatively little overlap, since they also require lingual-lingual coordination. Back-to-front coronal-labial sequences (e.g., [tp db])—like front-to-back labial-coronal sequences (e.g., [pt bd])—should then allow for more overlap because the tongue tip and lips are free to move into their constrictions at the same time.

In support of this biomechanical explanation, research on lingual motion into and out of dorsal gestures suggests that there are specific motor constraints that restrict the physical movement of the tongue. In their EMMA study on the movement of the tongue during German velar consonants [k ɡ ŋ], Mooshammer *et al.* (1995) observed that lingual movement is restricted to a motion that takes a clockwise, elliptical path, beginning with an upward movement into the dorsal constriction then transitioning into a forward and/or downward movement out of the constriction. This phenomenon, which they describe as "looping", was particularly strong when the following vowel was high and front. If tongue movement during dorsal gestures is constrained by the path that the tongue must take to move into and out of the constriction, then dorsal-coronal CC sequences are expected to involve substantial lag between the dorsal and coronal constrictions; the tongue must follow a posterior-to-anterior path rather than move into dorsal and coronal constriction locations simultaneously.

In the following sections, I describe the findings from both production and perception work that show evidence of an effect of greater overlap in front-to-back and back-to-front sequences. The term "place order" henceforth refers to the order of the place of the articulators in the CC sequence and additionally, in studies in which C2 is held constant as a coronal (i.e., "front-to-back"=[pt pl pɾ] and "back-to-front"=[kt kl kɾ]), the degree of physical independence between the articulators.

1.4.1 Production findings

A substantial body of information reveals that gestural overlap is systematically influenced by the order of the location (place) of the constriction during the production of CC sequences. Examining electropalatographic (EPG) data on [t#k] and [k#t] clusters in English, Hardcastle and Roach (1979) reported a shorter interval between the initiation of the closures for [t] and [k] in [t#k] than in [k#t] and hypothesized that this pattern was due to differences in the speed at which the tongue tip and the tongue dorsum move into place to produce a plosive constriction. That is, because the motion of the tongue dorsum is slower than that of the tongue tip, the interval between the closures for C1 and C2 is longer when C1 involves a tongue-dorsum gesture.

Two other EPG studies of lingual (i.e., coronal and dorsal) plosive-plosive sequences also yielded findings clearly indicative of a place-order effect. Byrd (1996) found longer intervals of C1-C2 palate-contact overlap during English heteromorphemic front-to-back [d#g] (*ba**d** **g**ab*) than during back-to-front [g#d] (*ba**g** **d**ab*). Peng (1996) found the same result for Taiwanese sequences involving heteromorphemic [t#k t#kʰ] (e.g., [pa**t**^{53} **k**aŋ^{55}] ‘another day’) and [k#t k#tʰ] (e.g., [la**k**^{53} **t**aŋ^{55}] ‘six years’). Results of acoustic studies are consistent with these findings, as shown for heteromorphemic CC sequences in English (Zsiga, 1994), word-initial CC clusters in Tsou (Wright, 1996), and word-initial and -medial CC clusters in Georgian (Chitoran, 1999).

More recently, Chitoran *et al.* (2002) used EMMA to track the movements of the tongue tip (TT), tongue dorsum (TD), and upper and lower lips (UL and LL) during the production of plosive-plosive sequences [bg pʰtʰ dg gb tʰb gd] by two native speakers of Georgian. They measured coronal, velar, and labial aperture values during the productions and, based on the aperture values, identified the time points at which C2 gestural onsets occurred relative to the onsets of C1 gestural constriction plateaus in order to calculate “gestural lag” values. In accordance with the expected place-order effect, both speakers of Georgian produced more gestural overlap in front-to-back sequences than in back-to-front sequences, although one speaker's productions showed this pattern only in word-initial position and not word-medially. Chitoran *et al.* interpreted their results as evidence that listeners' perceptual needs shape the spatiotemporal patterning of gestures produced by speakers, proposing that speakers avoid substantial intergestural overlap in back-to-front plosive-plosive sequences because release of a front constriction during a back closure is audible, whereas a front constriction would mask a back release.

In a follow-up study, Chitoran and Goldstein (2006) investigated the perceptual recoverability claim by testing word-initial Georgian sequences in which the two consonants have unequal constrictions, namely, plosive-liquid sequences ([**p'l**aʒ-i] 'beach' and [**br**az-i] 'anger' versus [**k'l**antʃ'-i] 'claw' and [**k'r**av-i] 'lamb') and trill-plosive sequences ([**rb**ev-a] 'to raid' and [**rk'**al-i] 'arc'). Instead of measuring temporal lag as the percentage of C1 constriction at which the C2 onset occurred, they measured lag as the duration of the interval between C1 and C2 gestural onsets. Their data appear to be inconsistent with the perceptual-recoverability hypothesis, as they found greater labial-coronal than dorsal-coronal overlap in these plosive-liquid and trill-plosive sequences, despite these sequences' being less likely to involve the degree of masking expected for overlapped plosive-plosive sequences. As a result, Chitoran and Goldstein (2006) offered a new interpretation of both the new and prior (Chitoran *et al.*, 2002) findings, arguing that the intergestural timing patterns are a potential result of a Georgian-specific pattern of articulatory coordination that has been phonologized and generalized to the Georgian grammar. Chitoran and Cohn (2009) revisited the 2006 data and reported not only a high degree of temporal separation between gestures in plosive-plosive sequences overall, but also a higher probability of vowel epenthesis between gestures in back-to-front than front-to-back plosive- plosive sequences. They speculated that vowel epenthesis in back-to-front plosive-plosive clusters is a further step towards the phonologization of the order-of-place timing effect in Georgian, although they did not specify whether this production phenomenon has a perceptual motivation.

Kühnert *et al.* (2006) presented findings consistent with Chitoran and Goldstein's Georgian data. In their EMMA study of the production of French words containing initial obstruent-lateral [pl fl kl] and obstruent-nasal [pn fn kn][1] sequences produced by two native speakers, Kühnert *et al.* measured: 1) the amount of overlap between the constriction interval of C1 and the entire gesture interval of C2 starting at its onset ("onset overlap"), and 2) the amount of lag between the end of the constriction interval of C1 and the beginning of the constriction interval of C2 ("constriction lag"). Holding C2 place to be coronal, they found consistently greater onset overlap and shorter constriction lag in front-to-back sequences [pl pn] than in back-to-front sequences [kl kn]. Kühnert *et al.* argued against a perceptual-recoverability account, given that acoustic masking effects due to place order should only influence the production of plosive-plosive sequences, but not that of plosive-nasal or plosive-lateral sequences.

1.4.2 Perception findings

The perceptual literature on the effect of place order on gestural overlap in CC sequences is limited to a few studies. Byrd (1992) used an articulatory synthesizer to create four continua for the English stimuli [bæ**b**#**b**æn] (*bab ban*), [bæ**b**#**d**æn] (*bab Dan*), [bæ**d**#**b**æn] (*bad ban*), and [bæ**d**#**d**æn] (*bad Dan*) that varied in percent gestural overlap and tested the extent to which listeners perceptually assimilated C1 to C2 as a function of gestural overlap. Using a two-way

[1] For more discussion on French obstruent-nasal sequences, as investigated in Kühnert *et al.* (2006), see Sections §1.5 and §1.6.

(/b/ or /d/), forced-choice identification paradigm, Byrd found that, as the percentage of gestural overlap increased in the front-to-back sequence [bæ**b**#**d**æn], C1 ([b]) was accurately identified (at least 80% of the time) until there was 100% temporal overlap between [b] and [d], at which point C1-identifications fell to around 60% accuracy. For the back-to-front sequence [bæ**d**#**b**æn], increasing gestural overlap inhibited successful perception of C1 (/d/) such that C1 responses fell to 40% accuracy at 100% gestural overlap between [d] and [b]. This finding is consistent with the perceptual-recoverability hypothesis: increased gestural overlap resulted in more assimilatory responses in the back-to-front sequence [d#b] than in the front-to-back sequence [b#d].

In a perceptual follow-up to her EPG production study on Taiwanese hetero-morphemic plosive-plosive sequences (CV**C1**#**C2**VC), which showed greater overlap for coronal-dorsal than for dorsal-coronal clusters, Peng (1996) tested the perceptual assimilation of C1 plosive codas. Peng excised the interval from the beginning of the first syllable up to the release of the plosive closure in the onset of the second syllable (C2) and presented these excised sequences to native-speaking Taiwanese listeners in an oddball paradigm, in which the task in each trial was to decide whether the identity of C1 (coda of the first syllable) in the stimulus matched the identity of the target C1 in the listening block. She found that coronal-dorsal sequences [t#k] and [t#k^h] were perceived as /k#k/ and /k#k^h/ more frequently than the dorsal-coronal sequences [k#t] and [k#t^h] were perceived as /t#t/ and /t#t^h/, indicating that the more overlapped, coronal-dorsal C1s were more prone to perceptual assimilation. This finding is inconsistent with the perceptual-recoverability account, since temporal overlap between C1 and C2 constrictions should have a higher likelihood of masking acoustic release cues in back-to-front (dorsal-coronal) than front-to-back (coronal-dorsal) sequences.

Surprenant and Goldstein (1998) assessed the relationship between the amount of gestural overlap and listeners' ability to detect C1 in English [C#C] sequences using a phoneme-monitoring task for C1. They excised the monosyllables [tat] *tot* and [tap] *top* from recordings of naturally produced utterances in the frame *my to*[C1#C2]*uddles*, where C2 was either [p] or [t], and tested whether there was a correlation between the amount of articulatory (C1-C2) overlap in each token (measured via x-ray microbeam imaging) and the accuracy of detection of C1 for that token. For the [t#p] utterances, they found that the more the tongue-tip and labial gestures overlapped, the perceptually less detectable C1 ([t]) was. However, although this effect was statistically significant when overlap was defined as the temporal lag between the release of the C1 and the release of C2 (i.e., "opening lag"), the correlation was not significant when defined as the temporal lag between the achievement of C1 and the achievement of C2 (i.e., "closing lag"). For the [p#t] utterances, they found no correlation between degree of overlap (both "opening" and "closing") and C1 perceptual detectability. This asymmetry in the relationship between CC overlap and detectability for coda [t] and for coda [p] may indicate that the release of a C1 *back* closure ([t]) is acoustically hidden when it occurs during the achievement of a C2 *front* closure ([p]), whereas the release of a C1 *front* closure ([p]) is audible when it occurs during the achievement of a C2 *back* closure ([t]).

Findings from two of the three perceptual studies on the recoverability of specific CC sequences are consistent with a perceptual-recoverability account of the place-order effect on intergestural overlap. Nonetheless, whereas all three of these studies show that perception is affected by overlap differences according to place order, they do not by themselves establish that perception itself was the source of the overlap pattern.

1.5 Effect of C1 Manner on CC Overlap

Biomechanics and perceptual recoverability make conflicting predictions as to how C1 manner influences gestural coordination in CC sequences. In a biomechanical account, the amount of gestural overlap in such sequences should, broadly speaking, depend relatively little on whether C1 is a plosive or a fricative if the articulators for C1 and C2 are held constant. In that approach, the specific timing patterns that are found among different CC sequences are determined by the physical constraints relevant to the coordination of two gestures, such as conflicting demands on the movement of the tongue body during co-occurring anterior (coronal) and posterior (dorsal) gestures. If the gesture for C1 is changed from a plosive to a fricative constriction, e.g., from [pt] to [ft], or from [kt] to [xt], the relationship between the articulators for C1 and C2 might remain very nearly the same. That is, we might expect that the relative demands on the lips, tongue tip and tongue dorsum present in [pt]~[kt] comparisons to also be present in [ft]~[xt] comparisons, even though there will likely be some alteration in the distance (and time) the articulator for C1 takes to reach its constriction target.

While biomechanics predicts no or little effect of C1 manner, perceptual recoverability predicts that a CC sequence with a fricative C1 should allow for more gestural overlap than a CC sequence with a plosive C1 if the cluster is to remain perceptible to listeners. Wright (1996), for example, argues that listeners have differing acoustic needs when attending to plosives and fricatives. Full cues to the identity of a fricative are available immediately when the aperiodic noise associated with frication assumes a spectral profile characteristic of that fricative, presumably at the time at which the fricative constriction is completed. While partially present in the F2 transition before constriction, acoustic cues to plosives encompass the release burst, which has been shown to play a role in the perception of the place of plosive constrictions (Liberman, 1954; Dorman *et al.*, 1977; Kewley-Port, 1983). If a plosive C1 needs to be coordinated with another plosive C2, then the constriction of C2 must be delayed until after the release of C1 if the complete set of acoustic cues to C1 is to be perceived by the listener.

Evidence for an effect of C1 manner of articulation on gestural overlap is sparser than that supporting the effects of word position and place order. EPG contact-profile data from Byrd (1996) show that the amount of C1-C2 overlap during the fricative-plosive sequence [s#g] (*ba**ss** **g**ab*) was greater than the overlap for the plosive-fricative sequence [g#s] (*ba**g** **s**ab*), as produced by native speakers of American English. However, Byrd's comparison involved CC pairs that contain a difference in both manner and place order, making it difficult to determine whether differences in overlap were due to both manner and place order versus manner alone.

Kühnert *et al.*'s (2006) EMMA study of French initial CC sequences provided evidence not only of place but also manner effects on overlap patterns: fricative-sonorant sequences [fl fn] were produced with longer onset overlap and shorter constriction lag than plosive-sonorant sequences [pl pn kl kn]. Kühnert *et al.*'s findings are consistent with those of Byrd (1996), even though the two studies examine different CC sequences, and Kühnert *et al.* offer a perceptual reason for the effect, suggesting that fricative [f] might tolerate more coarticulation with a following nasal or lateral than plosive [p k] due to the continuous noise (i.e., temporally extensive cues for C1) associated with frication for [f].

1.6 Effect of C2 Manner on CC Overlap

Two studies that investigate the role of C2 manner on the coordination of gestures in CC sequences generally find that plosive-plosive sequences are produced with the least intergestural overlap, while other sequence types, such as plosive-liquid and liquid-plosive, show greater intergestural overlap (Kühnert *et al.*, 2006; Chitoran & Cohn, 2009). Cohn and Chitoran (2009) revisited the Georgian data presented by Chitoran and Goldstein (2006) and observed a substantially greater degree of temporal separation (onset lag) between C1 and C2 in plosive-plosive sequences than in plosive-liquid sequences. This effect held in both back-to-front and front-to-back contexts. Additionally, they reported that their speakers occasionally epenthesize an interconsonantal vowel during plosive-plosive sequences, regardless of place order. By contrast, such vowels are never epenthesized into plosive-liquid sequences.

Kühnert *et al.* (2006) also found differences in gestural overlap as a consequence of C2 manner: French obstruent-lateral [pl fl kl] sequences were produced with greater onset overlap and shorter constriction lag than obstruent-nasal [pn fn kn] sequences. Their explanation for substantially less overlap in obstruent-nasal sequences was that nasal stops require the coordination of both an oral and velic closure and, since the articulation of a stop C1 requires sufficient intra-oral pressure build-up and hence a delay in nasal venting for the following nasal C2, both oral and velic closure gestures for C2 are unlikely to begin by the time a C1 stop closure is released. The fact that the effect was observed in both place orders points toward biomechanical rather than perceptual-recoverability factors. A caveat of Kühnert *et al.*'s (2006) result is the fact that words containing initial obstruent-nasal clusters [pn fn kn] were low-frequency loanwords and acronyms, and thus may have been subject to effects of lexical frequency and phonotactic violations in production.

Along perceptual-recoverability lines, an influence of C2 manner on gestural overlap is expected, but only or at least primarily in a back-to-front place order. Inasmuch as front-to-back and back-to-front sequences have been shown to vary in their perceptibility (refer to Section §1.4.2), substantial acoustic masking of C1 cues due to overlap is mainly expected when the CC sequence is back-to-front and plosive-plosive. In front-to-back plosive-plosive sequences, the release should be audible even when the constriction of C2 precedes the release of C1, and thus C1 and C2 should be able to overlap whether C2 is a plosive, a fricative, or liquid, without substantially degrading the quality of acoustic cues associated with C1. In back-to-front sequences, overlap should be avoided primarily when both C1 and C2 are plosives. If C2 is a fricative or a liquid (i.e., there is no complete closure), the release of a posterior C1 constriction even during a C2 constriction should be audible to at least some degree. If C2 is a plosive, then speakers are expected to lengthen the lag between C1 release and C2 constriction, so that the release of a plosive C1 is perceptible.

The biomechanical account—under which the timing of the C2 gesture depends on the spatiotemporal dynamics required to make that gesture—is only weakly supported by the effect of C2 manner shown in the literature. Consistent with the discussion in Section §1.5, the differences in the timing of the gestures for C1-C2 should be relatively small across manner (plosive, fricative, liquid) variations in C2, as long as the location of the constriction is held constant. Thus, timing differences according to manner arise primarily from minor differences in distance the articulator for the C2 gesture must travel to reach its target, and possible different requirements for target precision. In general, the effect of C2 manner on gestural coordination should be the same in both labial-coronal (front-to-back) and dorsal-coronal (back-to-front) sequences unless there are biomechanical constraints that influence the time course of articulator movement in one place order and not the other. If, for example, a particular language or speaker produces coronal plosive ([t]) constrictions with a more front lingual position than that for fricative ([s]) constrictions, then the tongue could take longer to coordinate a coronal stop with a dorsal stop ([kt]) than a more posterior coronal fricative with the same dorsal stop ([ks]). This pattern would arise only when the gestures for C1 and C2 are dependent on the same articulator (the tongue). In front-to-back CC sequences, if these were to involve the coordination of labial and coronal gestures, possible differences in constriction location for coronal plosives versus coronal fricatives should not matter, since the labial and lingual articulators are free to move independently of each other. This biomechanical account of the possible interaction between C2 manner and place order is somewhat tentative, since it relies on the characteristics specific to a particular language or speaker for which the exact location of coronal constrictions along the palate differ by manner.

1.7 Summary of Phonetic Effects on CC Coordination

Tables 1.1 and 1.2 summarize the production and perception studies, respectively, discussed in Sections §1.3-1.6, with results that support the biomechanical and/or the perceptual recoverability approach to gestural coordination indicated in the rightmost columns.

Effect	*Study*	*CC stimuli*	*Measurement*	*Effect observation*	*Which theory is supported?*	
					Biomechanics	*Perceptual Recoverability*
Within-word position	Wright (1996)	Tsou [pt pk tp tk pɗ tɓ kɗ]	Rate of plosive release	Lower rate of C1 release word-internally than word-initially.		✓
	Chitoran (1999)	Georgian [p^{h}t^h t^{h}b dg gd t^{h}k^h k^{h}t^h]	C1-C2 inter-burst interval (IBI)	Shorter IBI word-internally than word-initially.		✓
	Chitoran & Goldstein (2006)	Georgian [p^{h}l p'l p'r br k'l k'r rk' rk^h rb]	EMMA: Lag between C1 and C2 onset	Partial evidence for shorter lag time word-internally than word-initially.		✓
Place order	Zsiga (1994)	English [d#k] vs. [d#p]	F2 and F3 transition during V into C1	Greater influence by C2 on F2 and F3 into C1 in front-to-back than in back-to-front.	✓	✓
	Chitoran (1999)	Georgian [dg t^{h}k^h] vs. [gd k^{h}t^h]	C1-C2 inter-burst interval (IBI)	Shorter IBI in front-to-back than in back-to-front.	✓	✓
	Byrd (1996)	English [d#g] vs. [g#d]	EPG: Linguo-palatal contact profiles	Greater C1-C2 palate-contact overlap in front-to-back than in back-to-front.	✓	✓
	Peng (1996)	Taiwanese [t#k], [t#k^h] vs. [k#t], [k#t^h]	EPG: Linguo-palatal contact profiles	Greater palate-contact overlap in front-to-back than in back-to-front.	✓	✓

Effect	*Study*	*CC stimuli*	*Measurement*	*Effect observation*	*Which theory is supported?* Biomechanics	*Perceptual Recoverability*
Place order (cont.)	Chitoran *et al.* (2002)	Georgian [bg p^ht^h dg] vs. [gb t^hb gd]	EMMA: Onset overlap and constriction lag (see Section §1.4.1)	Shorter lag time in front-to-back than in back-to-front.	✓	✓
	Chitoran & Goldstein (2006)	Georgian [p^hl p'l p'r br rk' rk^h] vs. [k'l k'r rb]	EMMA: Lag between C1 onset and C2 onset	Shorter onset lag time in front-to-back than in back-to-front (as with Georgian plosive-plosive sequences).	✓?	✓?
	Kühnert *et al.* (2006)	French [pl pn fl fn] vs. [kl kn]	EMMA: Lag between release of C1 and the onset of C2	Greater onset overlap and shorter constriction lag in front-to-back than in back-to-front.	✓	✓?
C1 manner	Byrd (1996)	English [s#g] vs. [g#s]	EPG: linguo-palatal contact profiles	Greater C1-C2 palate-contact overlap in fricative-plosive than in plosive-fricative.		✓
	Kühnert *et al.* (2006)	French [pl pn kl kn] vs. [fl fn]	EMMA: Onset overlap and constriction lag (see Section §1.4.1)	Greater onset overlap and shorter constriction lag in fricative-C than in plosive-C.		✓
C2 manner	Chitoran & Cohn (2009)	Georgian [bg p^ht^h dg gb t^hb gd] vs. [p'l p'l p'r br k'l k'r]	EMMA: lag between onsets of C1 and C2 gestures	Greater C1-C2 overlap in plosive-liquid than in plosive-plosive.	✓?	✓
	Kühnert *et al.* (2006)	French [pl kl fl] vs. [pn kn fn]	EMMA: Onset overlap and constriction lag (see Section §1.4.1)	Greater onset overlap and shorter constriction lag in plosive-lateral than in plosive-nasal.	✓?	

Table 1.1. Summary of findings from speech production studies on intergestural timing effects during CC sequences. Checkmarks (✓) indicate which theory is supported, and weak evidence for a particular theory is indicated with a question mark (?).

Effect	*Study*	*CC stimuli*	*Measurement*	*Effect observation*	*Which theory is supported?* *Biomechanics*	*Perceptual Recoverability*
Place order	Byrd (1992)	English [d#b] vs. [b#d] (synthesized)	Rate of C1 assimilation to C2 place	More perceptual assimilation (C1-to-C2) as gestural overlap was increased for back-to-front than for front-to-back.		✓?
	Peng (1996)	Taiwanese [t#k], [t#k^h] vs. [k#t], [k#t^h] (excised from disyllables)	Rate of C1 assimilation to C2 place	More perceptual assimilation (C1→C2) for front-to-back (greater-overlap condition) than for back-to-front (less-overlap condition).		
	Surprenant & Goldstein (1998)	English [t#p] vs. [p#t] (naturally-produced, excised)	Accuracy of C1 detection	Greater influence of overlap on rate of C1 detection for back-to-front than for front-to-back.		✓?

Table 1.2. Summary of findings from speech perception studies on intergestural timing effects during CC sequences. Checkmarks (✓) indicate which theory is supported, and weak evidence for a particular theory is indicated with a question mark (?).

As is evident in Table 1.1, data on the effect of within-word position, i.e., greater gestural overlap word-medially than word-initially, supports a perceptual-recoverability account more strongly than a biomechanical one. If gestural overlap influences acoustics such that CC sequences are rendered less recoverable, then reduced gestural overlap in word-initial contexts may facilitate lexical access by listeners.

Most findings on an effect of place order—i.e., greater overlap in front-to-back than in back-to-front—supports both perceptual-recoverability and biomechanical hypotheses, and the uniformity of the effect across studies means that it is still unclear which of the two theories—if either—is better supported. Additional perceptual data in Table 1.2 only weakly support the perceptual-recoverability account, given that these data show that listeners' perceptual responses differ for the two place orders, but not that perception is the source of gestural timing differences according to place order produced by speakers.

The source of the C1-manner effect is also not well understood because the relevant findings exhibit conflicting patterns and thus support different theories. Given that the two relevant studies used different stimuli and tested different languages, additional testing on the influence of C1 manner on gestural coordination is needed so that the contributions of biomechanics and perception can be examined more closely.

In terms of the C2 manner, perceptual recoverability predicts an effect of C2 manner in back-to-front sequences, but not in front-to-back sequences. However, difficulty in teasing apart support for perceptual-recoverability and biomechanical accounts arises from the need, in the latter account, to have a detailed understanding of the possibly language- or speaker-specific articulatory requirements for the relevant gestures. If the articulatory demands on lingual-lingual coordination changes as C2 manner changes—i.e., from a plosive to a fricative or a lateral—the biomechanical account might also predict an effect of C2 manner, for example, in dorsal-coronal (back-to-front) [kt] vs. [ks] and not in labial-coronal (front-to-back) [pt] vs. [ps]. An investigation of the effect of C2 manner in both place orders is needed to further assess the theoretical claims.

1.8 Dissertation Design

This dissertation aims to resolve some of the indeterminacies among findings in the literature by exploring the interaction between the targeted linguistic effects discussed in Sections §1.4-1.7. By observing how changes in place order, C1 manner, and C2 manner influence how speakers time gestures in CC sequences, this study tests the claims that biomechanics and perceptual recoverability make about the gestural coordination.

Direct comparisons between the targeted effects on gestural timing are only possible in a language that permits the proper set of CC sequences. Standard Modern Greek is a suitable language for this type of study because it permits a set of CC sequences important to testing the theoretical predictions. The list of the Modern Greek CC sequences of interest is shown in Table 1.3, with CCs involved in critical comparisons (explained below) marked with asterisks.

	C1 = coronal		C2 = coronal	
	Front-to-back	Back-to-front	Front-to-back	Back-to-front
Plosive-plosive			pt*	kt*
Plosive-fricative			ps*	ks*
Plosive-liquid			pl* bl pɾ* bɾ	kl* gl kɾ* gɾ
Fricative-plosive	sk	sp	ft*	xt*
Fricative-fricative	sx	sf zv	fθ vð	xθ ɣð
Fricative-liquid			fl vl fɾ vɾ	xl ɣl xɾ ɣɾ

Table 1.3. A (partial) set of CC sequences that occur both word-initially and word-medially in Modern Greek, sorted by C1-C2 *manner* (rows) and *order of place* (paired columns). Sequences marked with an asterisk belong to critical comparisons for the four targeted timing effects.

Because a study that includes intergestural timing comparisons between all of the CC sequences in Table 1.3 would become impractically large, this study focuses on only a smaller set of critical Modern Greek CC contrasts. Table 1.4 lists the four targeted CC timing effects found in the literature, their contrastive conditions, and the CC sequences of Modern Greek that pertain to these contrasts. Not only do these CC sequences occur in the desired contexts in Modern Greek, but they also occur relatively frequently in each context, and numerous morphologically unrelated Greek words are available for each condition. With this breadth of CC sequences in Greek, the effects of word position, place order, C1 manner, and C2 manner can be compared across Greek CC sequences, without introducing many confounding variables.

Effect	Conditions	CC sequences
Within-word position	word-initial	#CC…
	vs. word-medial	…CC…
Place order	front-to-back	[pt ps pl pɾ ft]
	vs. back-to-front	[kt ks kl kɾ xt]
C1 manner	plosive C1	[pt kt]
	vs. fricative C1	[ft xt]
C2 manner	plosive C2	[pt kt]
	vs. fricative C2	[ps ks]
	vs. liquid C2	[pl pɾ kl kɾ]

Table 1.4. Articulatory timing effects, representative CC types, and their corresponding, minimally-contrastive CC sequences occurring in Modern Greek.

This study focuses on just three of the four effects: place order, C1 manner, and C2 manner. For this dissertation, word-initial rather than word-medial CC sequences were analyzed because word onsets are more important than word-medial clusters to lexical access and word recognition (see Section §1.3) and are thus more ideal for testing whether and how speakers perceptually enhance their phonetic productions.

Overall, previous production studies on gestural timing in CC sequences fall short of differentiating the extent to which biomechanical and perceptual-recoverability approaches offer compelling accounts of the source of the articulatory timing effects of place order because they have not thoroughly investigated how the effects of C1 manner and C2 manner are affected by place order. Thus, this study on the gestural coordination of Modern Greek CC clusters should further current understanding of how biomechanical and perceptual-recoverability factors influence articulatory timing.

In Chapter II, I present the methodology of a speech production experiment that uses ultrasound and camera imaging and systematically manipulates place order, C1 manner, and C2 manner for Modern Greek CC sequences. Chapter III presents the results of the speech production experiment, and Chapter IV discusses the theoretical implications of these findings, taking into account the principal literature and hypotheses summarized in this chapter.

CHAPTER II
Methodology and Predictions

Ultrasound is an effective technique for imaging the surface of the tongue during speech production. Ultrasound images of large sections of the tongue surface can be used to analyze the dynamics of lingual movements involved in speech production. Previous work has studied, for example, the spatio-temporal characteristics of movements of the tongue tip toward the alveolar ridge/teeth and that of the tongue dorsum toward the velum (Wrench & Scobbie, 2003; Gick *et al.*, 2006; Benus & Gafos, 2007; Mielke *et al.*, 2011). Since ultrasound is both non-invasive (the ultrasound transducer remains completely external to the body) and unobtrusive (the transducer remains in a submental position), it is an excellent technology for investigating naturally-produced utterances involving rapidly articulated movements that, with other technologies, might be influenced by articulometric sensors positioned within the oral cavity.

In addition to an ultrasound system, which is able to image movements of the lingual articulators (i.e., tongue tip and dorsum), this study used a video micro-camera to track the movement of the upper and lower lips. This allowed for the investigation of labial gestures in consonants such as [p] and [f], which are crucial sounds in the comparison of contrasting place orders (back-to-front vs. front-to-back) and manners (plosive vs. fricative). Imaging data collected via the combined ultrasound and camera set-up provided information about the simultaneous location and movement of three supraglottal consonant articulators, that is, the lips, tongue tip, and tongue dorsum.

2.1 Design

2.1.1 Test stimuli

The set of test stimuli included five target words for each of the test conditions for the Modern Greek CC sequences listed in Table 1.3, where [Cl] and [Cɾ]-type clusters are considered to be members of the same category (plosive-liquid). This yielded a total of 16 conditions and a full list of 80 target words (= 8 cluster types × 2 word positions × 5 target words). All target words were selected from the *GreekLex* lexical-frequency corpus (Ktori *et al.*, 2008). Because the word-initial [kt] cluster (back-to-front, plosive-plosive) was found to occur strictly in pre-[i] contexts in Greek (with the exception of the low-frequency word *κτέρισμα* [ˈkte.ɾi.zma] 'grave offerings'), target words across conditions were restricted to pre-front vowel contexts to avoid dramatic kinematic differences between contexts that differ in quality of the following vowel. For each target cluster, the immediately preceding vowel (which was from the carrier phrase for the word-initial clusters and within the target word for medial clusters) was limited to [a] to ensure that the relevant articulator would be in a relatively open position at the onset of each CC-sequence production. Also, for each CC condition, the selected target words contained primary word stress on the vowel that immediately followed the CC sequence. Each word was two syllables when possible, although only three- or four-syllable words were possible for many of the word-medial CC conditions. Nearly all target words were nouns, verbs, or adjectives, and each word had a lexical frequency above zero as reported in *GreekLex*. Example words for each condition are given in Table 2.1. The set of 80 target stimuli was combined with 20 additional non-CC filler words, giving a total of 100 stimuli. A complete list of the target and filler words is provided in the Appendix, §A.1–A.3.

Each production session involved six iterations of each of 80 target words and 20 filler words, resulting in a total of 600 utterances (= 480 utterances containing target items + 120 utterances containing filler items). Speakers produced each word in the Greek carrier sentence: *Είπα* ___ *και πάλι* [ˈi.pa ____ ce ˈpa.li] (‘I said ____ again’). This carrier sentence was chosen to facilitate ultrasound- and camera-video synchronization with the audio signal (see Sections §2.3.1 and §2.3.2). The average duration of the actual recording time was approximately 40 minutes, although pre- and post-session instruction, as well as a ten-minute break in the middle of each recording session, lengthened the time that speakers spent in the recording room to roughly 75 minutes. Before each recording, participants reviewed the entire list of test and filler words and verified whether they were able to pronounce any words in the list that were unknown to them. While two to three of the most infrequent words were unknown to some speakers, all speakers reported their ability to pronounce these words when seeing it written.

	Labial-coronal		Dorsal-coronal	
	Word-initial	Word-medial	Word-initial	Word-medial
Plosive-plosive	[ˈ**pt**i.si] *πτήση* ‘flight’	[va.ˈ**pt**i.zo] *βαπτίζω* ‘I baptize’	[ˈ**kt**i.si] *κτήση* ‘possession’	[a.ˈ**kt**i] *ακτή* ‘coast’
Plosive-fricative	[ˈ**ps**i.fos] *ψήφος* ‘vote’	[ta.ˈ**ps**i] *ταψί* ‘pan’	[ˈ**ks**i.fos] *ξίφος* ‘sword’	[ta.ˈ**ks**i] *ταξί* ‘taxi’
Plosive-lateral	[ˈ**pl**i.o] *πλοίο* ‘ship’	[a.ˈ**pl**i] *απλή* ‘simple (FEM. SG.)’	[ˈ**kl**i.ma] *κλίμα* ‘climate’	[i.ɾa.ˈ**kl**is] *Ηρακλής* ‘Hercules’
Fricative-plosive	[ˈ**ft**i.no] *φτύνω* ‘(I) cough up’	[a.ˈ**ft**i] *αυτί* ‘ear’	[ˈ**xt**i.zo] *χτίζω* ‘(I) construct’	[a.ˈ**xt**i.ða] *αχτίδα* ‘ray’

Table 2.1: Example target words for each CC condition in the experiment.

2.1.2 Participants

Participants recruited for the production task were ten native speakers of Standard Modern Greek who were current undergraduate students, graduate students, or visiting faculty at the University of Michigan (average age: 28 years). However, due to issues related to the quality of two speakers' ultrasound recordings, only data from the remaining eight speakers were analyzed (see Section §3.1).

2.2 Data Collection

During the recording of each stimulus set, speakers were presented with five identical carrier phrases, each containing a target or filler item. This presentation was performed with the *Articulate Assistant Advanced* software (Wrench, Articulate Instruments, Queen Margaret University) (henceforth, AAA), in its "Record Ultrasound" mode.[1] Complete sets of recording phrases appeared at the top of the recording screen, as shown in Figure 2.1, with the entire screen visible to the experimenter and only the top portion visible to the speaker at all times throughout each session. During recording, the three data signals—audio, ultrasound, and camera—were collected as three independent streams, as described below.

2.2.1 Audio collection

Acoustic data were recorded with an AKG C400B condenser microphone at 44.1 kHz and streamed into a desktop Dell PC via a UA-25 USB audio interface (Roland Corporation). Audio file recording was initiated in AAA using the recording function in the "Record Ultrasound" mode.

[1] The flexibility of the "Record Ultrasound" mode in recording from a variety of video streams beyond ultrasound was utilized in order to obtain video of the lip movements with a camera (see Section §2.2.3), instead of video of lingual movements in ultrasound. For details on how ultrasound videos were streamed, refer to Section §2.2.2.

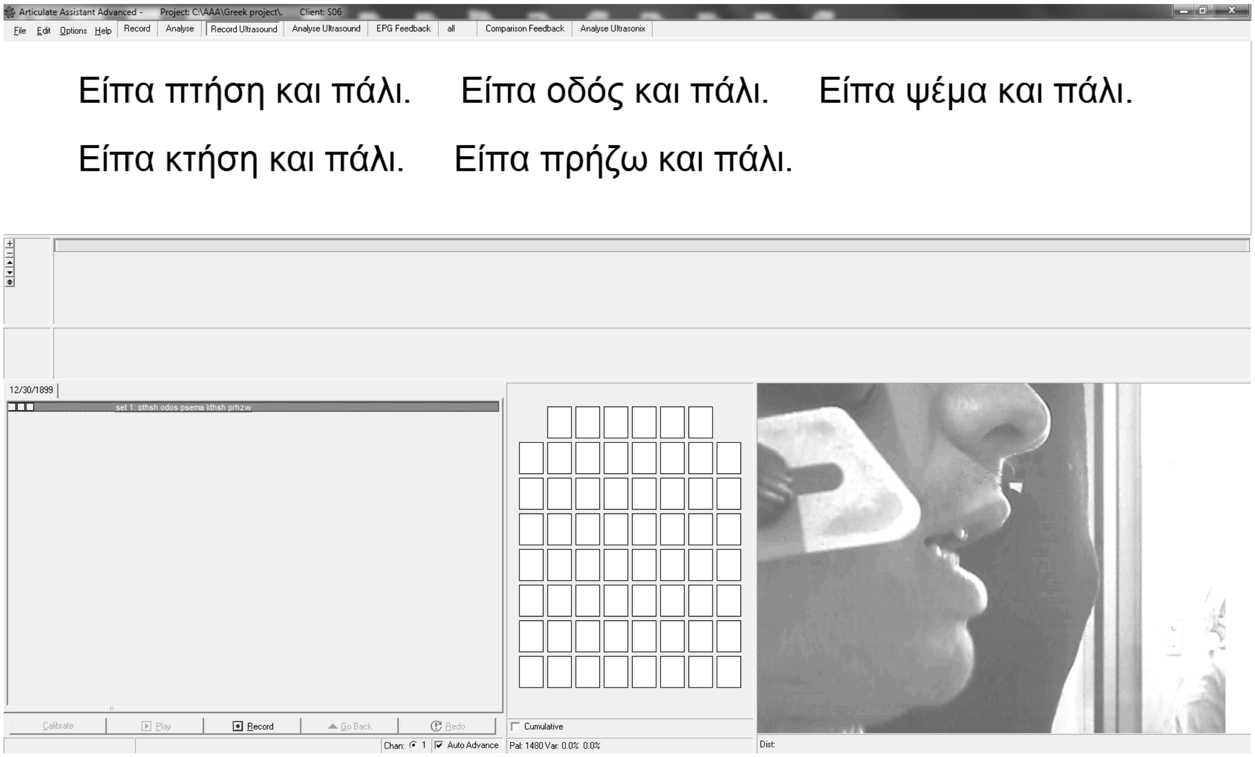


Figure 2.1: Example screen display during the recording of a five-item recording set. Although the experimenter saw the entire screen and thus could check for camera recording issues, only the top portion of the screen, which containing Greek text only, was visible to the speaker.

2.2.2 Ultrasound video collection

To determine the patterns of intergestural timing in each CC condition, lingual-movement data were collected using ultrasound midsagittal tongue imaging techniques. All ultrasound imaging data were collected in a sound-attenuated recording booth at the University of Michigan using a portable z.one *mini* ultrasound system (ZONARE Medical Systems, Inc.) and a P4-1c phased array transducer. Lingual movements were imaged at a frame rate of 60 frames per second (fps). Because consonantal gestures, particularly those corresponding to plosive closures and releases, are typically quick and short in duration, a rate of 60 fps was highly desirable. To accomplish this, the scanning procedure required streaming of ultrasound video directly onto an external medical digital recorder (MDR) at 9.5-second intervals and then burning of the individual video files into a *Digital Imaging and Communications in Medicine* (DICOM) format onto DVDs. These video files were subsequently converted into uncompressed AVI-format movies, which were imported into AAA for analysis.

During the collection of ultrasound videos in each experiment session, the ultrasound transducer was positioned underneath the jaw so that tongue-tip movement toward the alveolar ridge and teeth and tongue-dorsum movement toward the velum were maximally visible within the transducer's scanning field. The transducer was fixed in place with respect to the speaker's head using an Ultrasound Probe Stabilisation Headset (Articulate Instruments, Queen Margaret University), shown in Figure 2.2. Efforts were made to ensure that speakers did not wear the stabilization device for intervals longer than 20 to 30 minutes, and breaks were given to speakers halfway through each recording session (every 60 recording sets).

In order to establish a reference contour that corresponded to the shape and location of most of the palate, speakers were also instructed to sip and swallow a small bolus of water while monitoring the corresponding real-time ultrasound image on a computer screen. This procedure was repeated a total of ten times throughout each recording session, specifically, two times before the first set and two more times after every 30 recording sets.

Beeps emitted by the MDR at the beginning and end of the recording of each ultrasound movie file were used to achieve an initial, approximate synchronization of the ultrasound videos with the audio recordings post-collection. The acoustic beep for each recording set occurred around 50 milliseconds before the beginning of each 9.5-second ultrasound video file. Using this timing, 9.5-second audio intervals were extracted from the original audio recordings and dubbed to the silent ultrasound videos. The dubbed ultrasound videos were imported into AAA for analysis, at which point a more precise synchronization technique between the audio and ultrasound signal was applied (see Section §2.2.4).

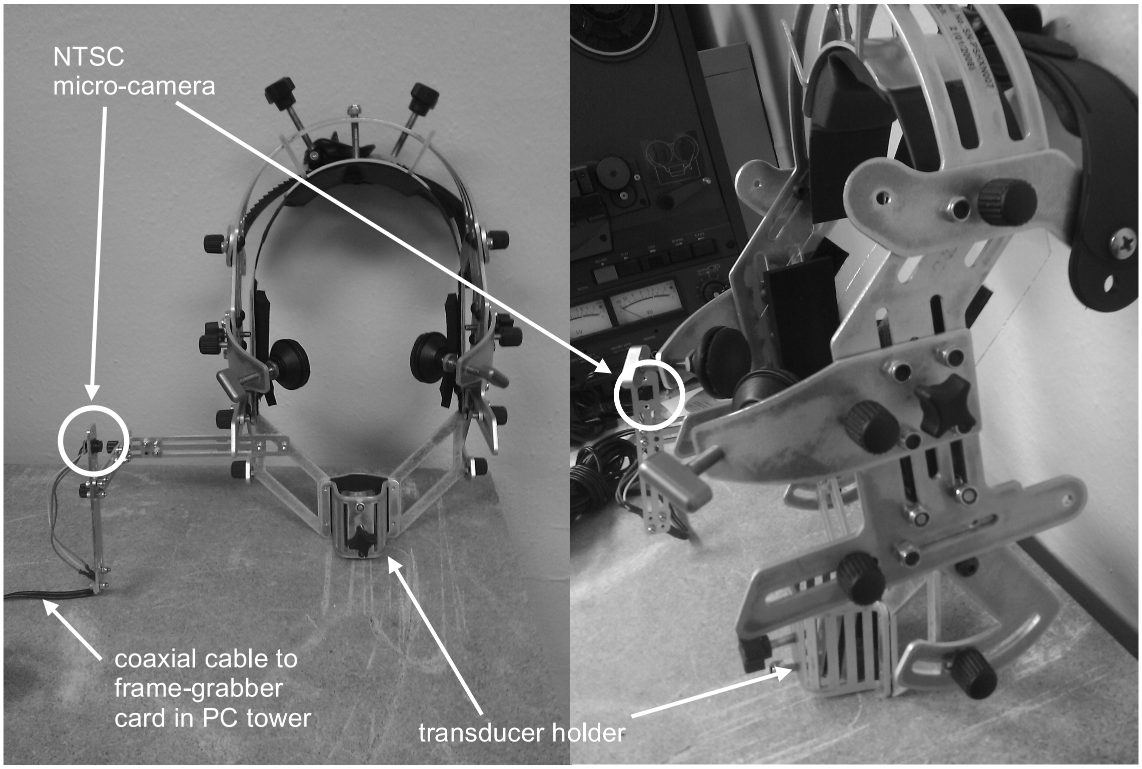


Figure 2.2: Front view (left) and left-side view (right) of the Ultrasound Stabilisation Helmet with side-arm attachment for micro-camera (discussed in Section §2.2.3).

2.2.3 Camera video collection

To determine the patterns of intergestural timing in front-to-back CC sequences, lip movement data were collected using a video micro-camera. This micro-camera, constructed by Alan Wrench (Queen Margaret University, 2012), was mounted with a lightweight side-joint onto the right side of the Ultrasound Stabilisation Helmet for concurrent ultrasound and lip-camera data collection. The camera captured frames according to the NTSC standard (30 fps interlaced). Because each frame consisted of two interlaced sets of scan lines captured at two different times, camera video frames were de-interlaced post-collection in order to yield non-interlaced images with half of the original visual resolution but presenting information from only a single time point. Thus, 60 fps is the actual frame rate of the camera video, since each set of the interlaced scan lines is captured at twice the speed as reported for interlaced scanning. The camera video was streamed onto the same computer that recorded the acoustic signal. Simultaneous video and audio collection was possible with the AAA software, which uses a frame-grabber PCI card to buffer video frames as they are transmitted from the NTSC camera. The synchronization of the camera video with the simultaneously recorded audio is explained in more detail in Section §2.2.4.

The video files from the micro-camera were saved in an uncompressed AVI format and were analyzed, along with audio and imported ultrasound video, in the AAA interface. The set-up of the computer, ultrasound, and MDR outside of the recording booth and the set-up of stimulus-presentation screen and microphone within the recording booth are shown in Figure 2.3.

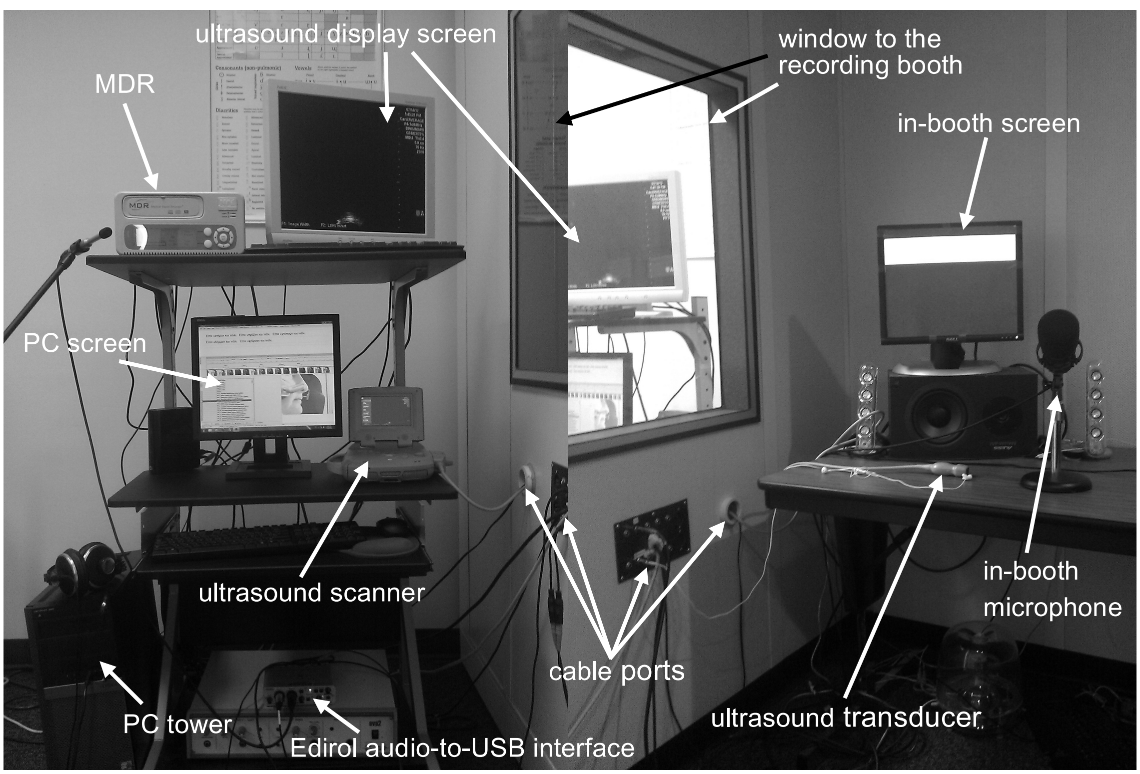


Figure 2.3: The set-up for the equipment used outside of the recording booth (left) and within the recording booth (right).

2.2.4 Synchronization

Precise synchronization between the data taken from the two video signals and time points in the audio signal per recording set was achieved by aligning articulatory events in the video with their corresponding acoustic landmarks. The use of plosive releases in the carrier phrase *Είπα ___ και πάλι.* [ˈi.pa ___ ce ˈpa.li] (‘I said ___ again.’) facilitated the identification of these events in the video streams and in the audio.

To synchronize ultrasound with the audio signal, articulatory releases of the palatal plosive [c] in *και* [ce] in the carrier phrase were identified. Articulatory release of [c] was defined as time of the last ultrasound frame showing a complete palatal closure position of the tongue dorsum during the entire [c]-closure movement. This method of locating the time of articulatory [c] release is shown in Figure 2.4. Because each recording set contained five sentences, there were (at least) five instances of the palatal stop release per ultrasound recording.

Corresponding acoustic times for the release of [c] in the carrier-phrase word *και* [ce] were identified as the onset of the release burst for [c] in the audio signal. Figure 2.5 shows the temporal alignment between the acoustic burst of [c]-release and the articulatory release of [c] as determined from ultrasound video. Finally, synchronization between ultrasound video and audio was achieved by aligning the articulatory onset of [c]-release (Figure 2.4) with the acoustic onset of the [c]-release in the acoustic recording (Figure 2.5). Because each recording set contained five [c] productions from the carrier phrase, a single recording set’s ultrasound video and corresponding audio were aligned according to the average offset between video and audio for the five [c]-releases in the set.

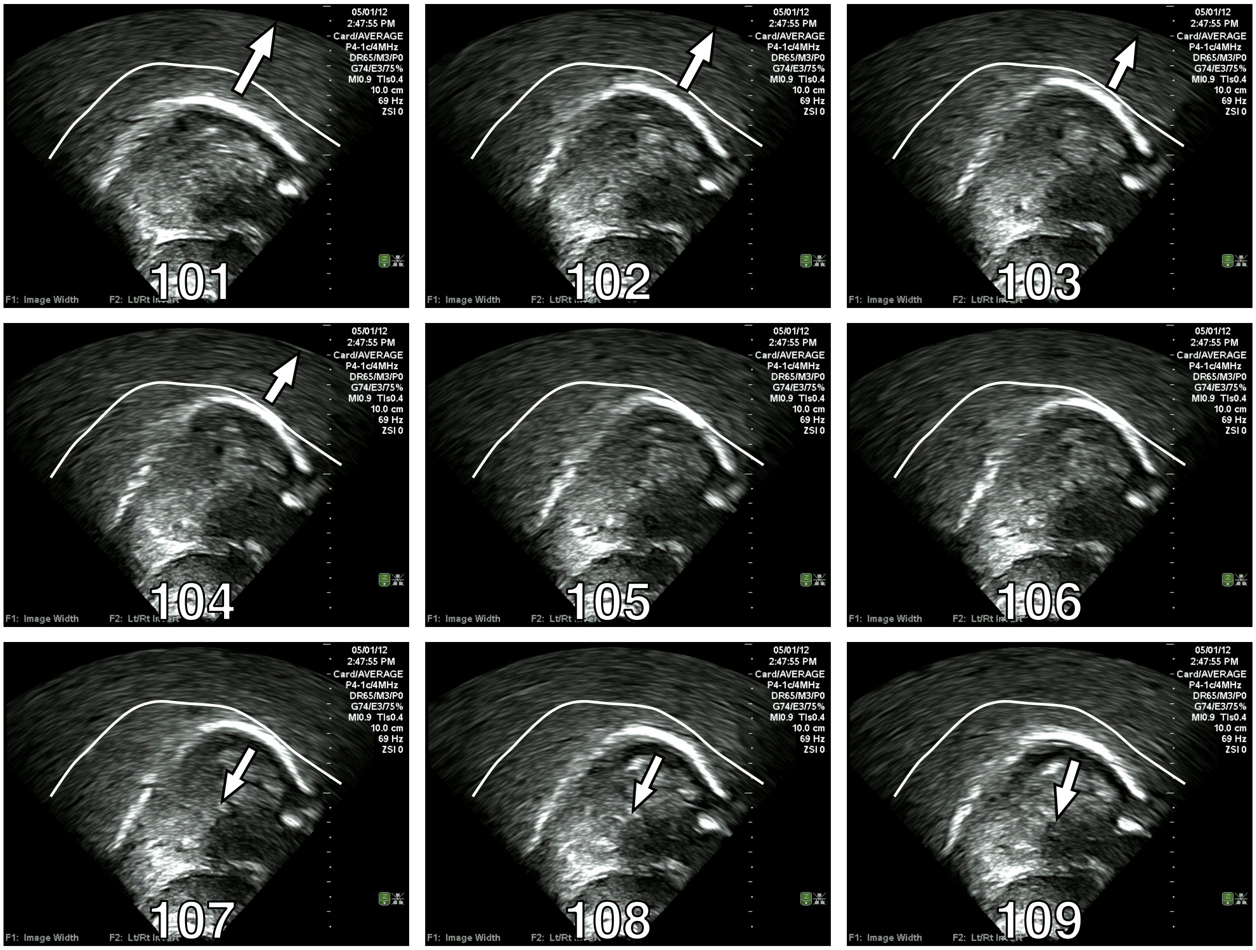


Figure 2.4: A series of nine ultrasound frames during [c] from *και* [ce] 'and' in the carrier phrase *Είπα* ___ *και πάλι* [ˈi.pa ___ ce ˈpa.li], as produced by speaker S06, showing the movement of the tongue dorsum towards closure (frames 101–103), during closure (frames 104–106), and out of closure (frames 107–109). The tongue tip is located to the right, and the tongue root is located to the left. The thin white curves above the tongue contour in each frame are identical traces of the speaker's palate for the portion of the session during which the pictured utterance was recorded. This palate trace was identified using scan sequences of multiple instances of swallowing of a water bolus by the speaker (as described in Section §2.2.2).

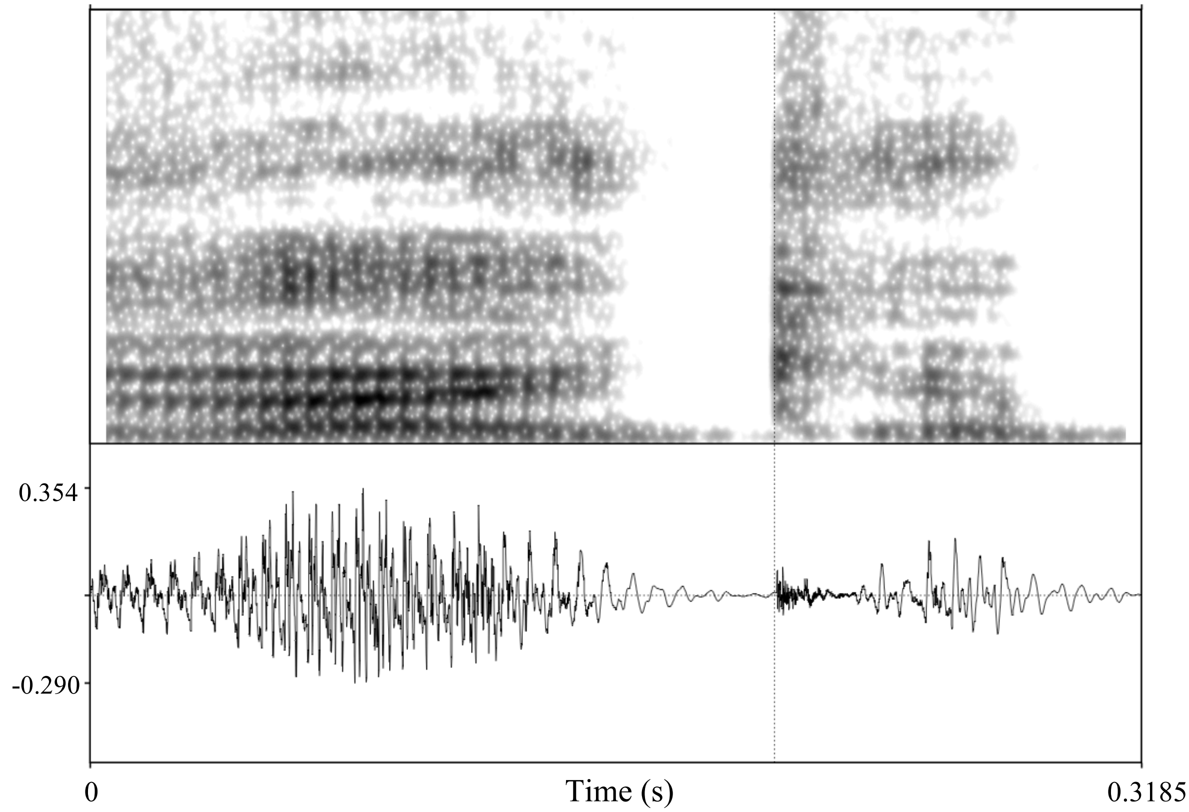


Figure 2.5: Identifying the acoustic correlate of the release of the palatal stop in *και* [ce] (sample taken from speaker S06). This is the same [c]-closure for which the ultrasound images are shown in Figure 2.4. The vertical dotted line marks the onset of the acoustic release burst, which occurs at the same time as frame 107 in Figure 2.4.

2.3 Data Analysis

2.3.1 Ultrasound video analysis techniques

Tongue-contour data were analyzed in AAA, which allows for smoothing splines to be drawn directly onto detectable boundaries within grayscale images that contain varying degrees of brightness. For ultrasound analysis, these smoothing splines were fitted along a radial grid in a “fan” shape across a given sequence of frames using at automated spline-fitting function, as shown in Figure 2.6.

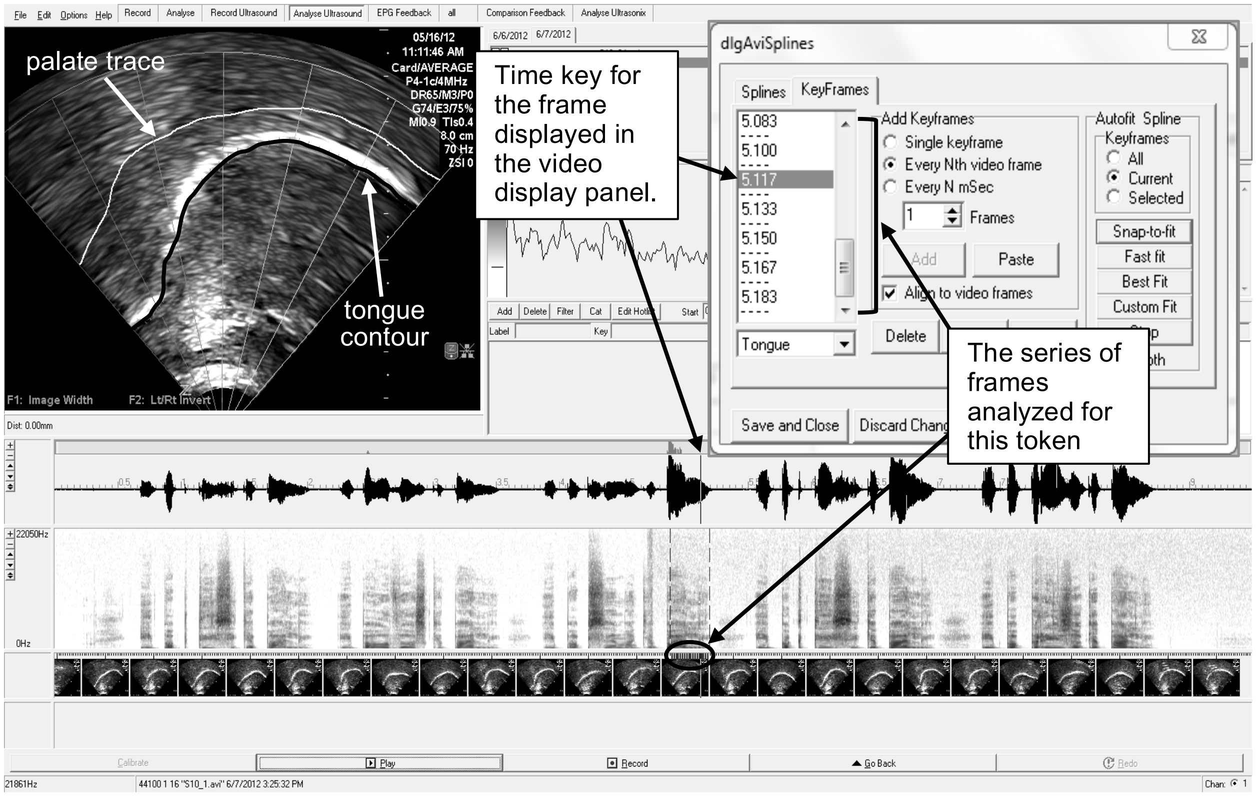


Figure 2.6: An example of the AAA interface used to analyze lingual contours in ultrasound video. The upper left panel shows an example of a tongue contour and palate trace drawn onto a single ultrasound video frame. The frame shown here occurs during the final [i] in the utterance *Είπα ψέμα και πάλι* [ˈi.pa ˈpse.ma ce ˈpa.l**i**], in the middle of a five-utterance set produced by speaker S10. Palate and tongue contours were fitted using a B-splines and "snakes" algorithm within the AAA software.

Since the number of axes on the radial grid was rather large (42 radii), the most important parts of the midsagittal cross-section of the tongue in the articulation of lingual consonants were not missed by this method of fitting splines to tongue-surface contours. Regions along the palate corresponding to each speaker's alveodental and velar regions of constriction were identified based on the locations at which coronal and dorsal constrictions, respectively, occurred in multiple, randomly selected tokens of dorsal and coronal constriction gestures among the items analyzed using ultrasound, as shown in Figure 2.7.

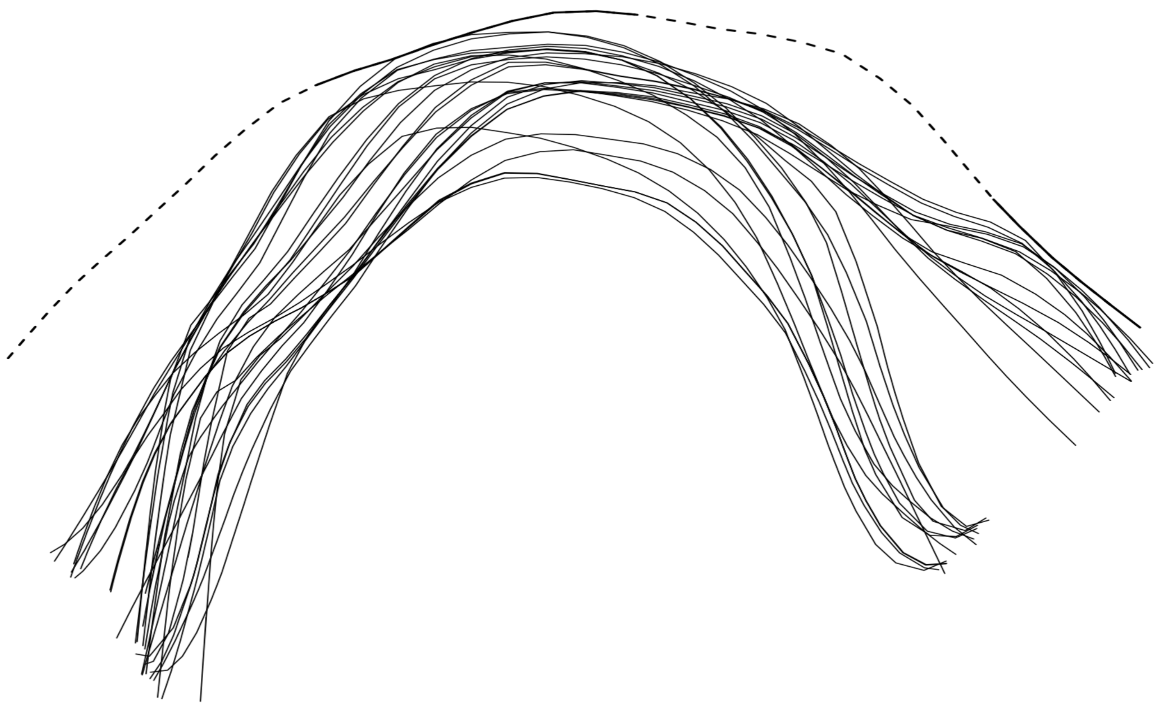


Figure 2.7: Example of aggregated tongue contours (overlapping series of solid lines) from frames from a single production of [kt] in *κτήση* [ˈ**kt**i.si] ‘possession’, produced by speaker S07. The top dashed line represents the trace of the palate, while the solid portions of this contour indicate the dorsal (left) and coronal (right) regions of constriction.

Apertures of the tongue tip (TT) in the alveodental region and of the tongue dorsum (TD) in the velar region were then calculated as the minimal distance (for all contour points) between the tongue contour within the TT- and TD-regions and the corresponding region on the palate trace. For each frame, the value for minimal aperture in each region was determined by finding the shortest distance between the points along the palate trace lying within the region and all spline coordinates along the tongue contour for that frame. Figure 2.8 shows a series of ultrasound frames during a production of a dorsal-coronal CC sequence, with increasing aperture in the TD region over the sequence and decreasing aperture in the TT region up to the fourth frame.

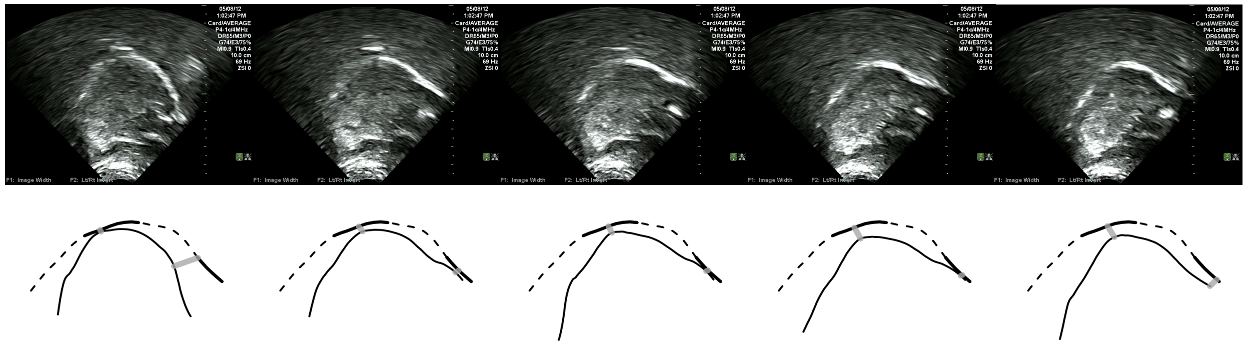


Figure 2.8: A sample series of ultrasound frames (top) out of the release of [k] and into and out of a constriction for [t] during the cluster [kt] in a token of *κτήση* [ˈ**kt**i.si] ‘possession’, produced by speaker S07. Below each frame is the corresponding plot of the tongue contour (completely solid lines) and palate trace (dashed lines) as drawn in AAA, with dorsal and coronal regions of the palate (solid bold sections within dashed lines) and location of minimal aperture in those regions (gray lines) indicated. For the purpose of showing the entire [k]-to-[t] motion, frames were taken 50-ms apart (20 fps), although analysis for this token was performed at a rate of 60 fps (16.7 ms between frames).

For each speaker, measures of aperture in the TT and TD regions were normalized to a scale from minimum to maximum aperture distance observed across the speakers’ CC productions. This value is reported in percentages, where zero percent indicates complete closure. A sample aperture plot for a single token of [kt] is presented in Figure 2.9.

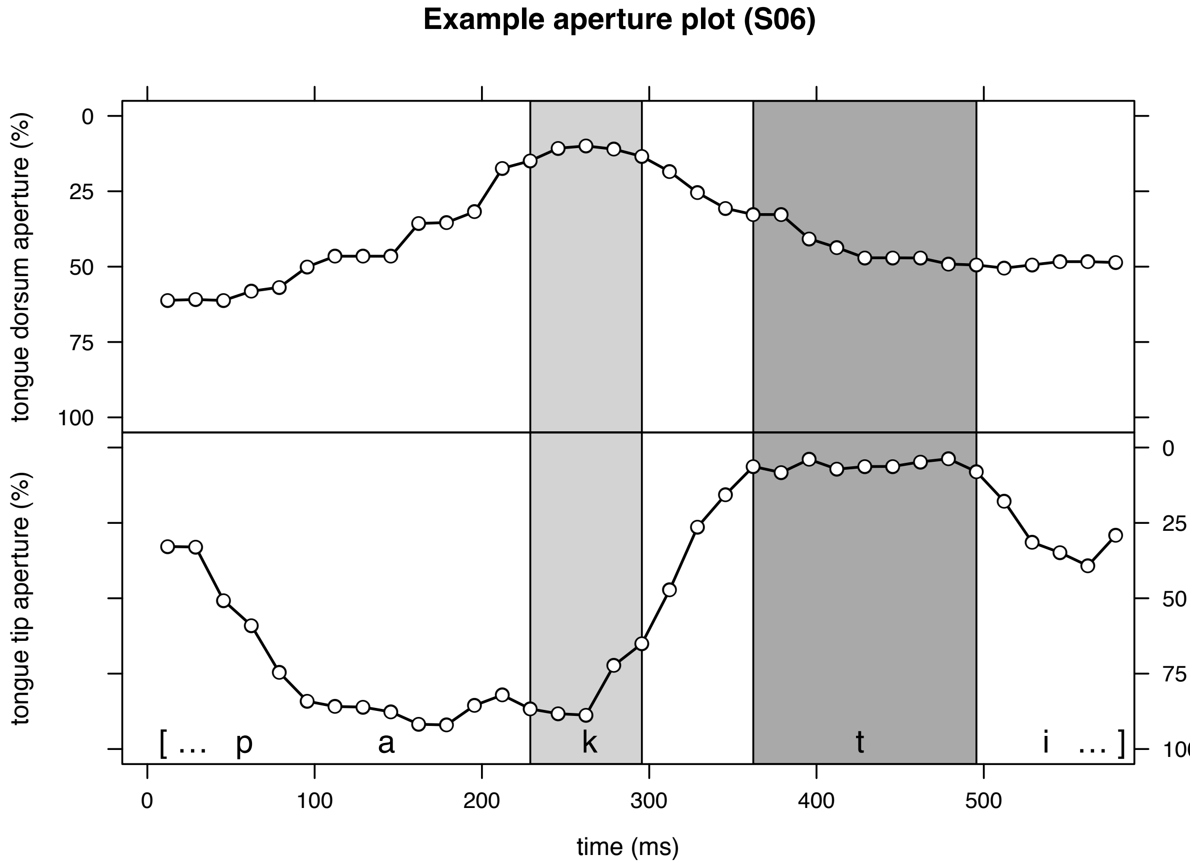


Figure 2.9: Tongue dorsum (top) and tongue tip (bottom) aperture, measured in percentage of total aperture range for each lingual articulator, against time, during the production of *(Eί)πα κτί(σμα)…* [(ˈi).pa ˈ**kt**i.(zma)…] by speaker S06. Vertical lines distinguish articulatorily defined boundaries for the closure intervals of the two segments [k] (light gray interval) and [t] (dark gray interval), discussed in Section §2.3.3.

2.3.2 Camera video analysis techniques

Similar to the method used for ultrasound video, synchronization between the data taken from lip-camera video frames and the corresponding audio stream was achieved by relating the articulatory release times of the bilabial plosive [p] in words *είπα* [ˈi.pa] and *πάλι* [ˈpa.li] in the carrier phrase *Είπα* ____ *και πάλι* [ˈi.pa ____ ce ˈpa.li] to the corresponding acoustic times of [p] release in audio signal. In the camera stream, the time of articulatory release of [p] was identified by examining the position of the upper and lower lips into, during, and out of labial closure. As shown in Figure 2.10, a single labial closure gesture involves the approximation of the lips toward each other until the opening between the lips is occluded. After complete closure, the lips may continue to move toward each other, resulting in further compression of lip tissue. Lip closure was defined as the interval during which there was no visible aperture between the upper and lower lips (frames 88 to 92 in Figure 2.10), and thus articulatory release of [p] was defined as occurring during the frame in which the frontmost portions of the upper and lower lips become spatially separate from each other (frame 93 in Figure 2.10). This separation often appears as a blurriness in the space between the lips caused by the quick movements of the articulators during such a frame.).

Corresponding acoustic times for the release of [p] in the carrier-phrase words *είπα* [ˈi.pa] and *πάλι* [ˈpa.li] were identified as the onset of the release burst for [p] in the audio signal. Figure 2.11 shows the temporal alignment between the acoustic burst of [p]-release and the articulatory release of [p] as determined from ultrasound video. Synchronization between lip camera video and audio was achieved by aligning the articulatory release of [p] (frame 93 in Figure 2.10) with the onset of the acoustic burst from [p]-release in the audio recording (the dotted line in Figure 2.11). Because each recording set contained five sentences, there were five instances of both words *είπα* [ˈi.pa] and *πάλι* [ˈpa.li] per recording file and thus a minimum of ten instances of the bilabial plosive [p] in each recording set.

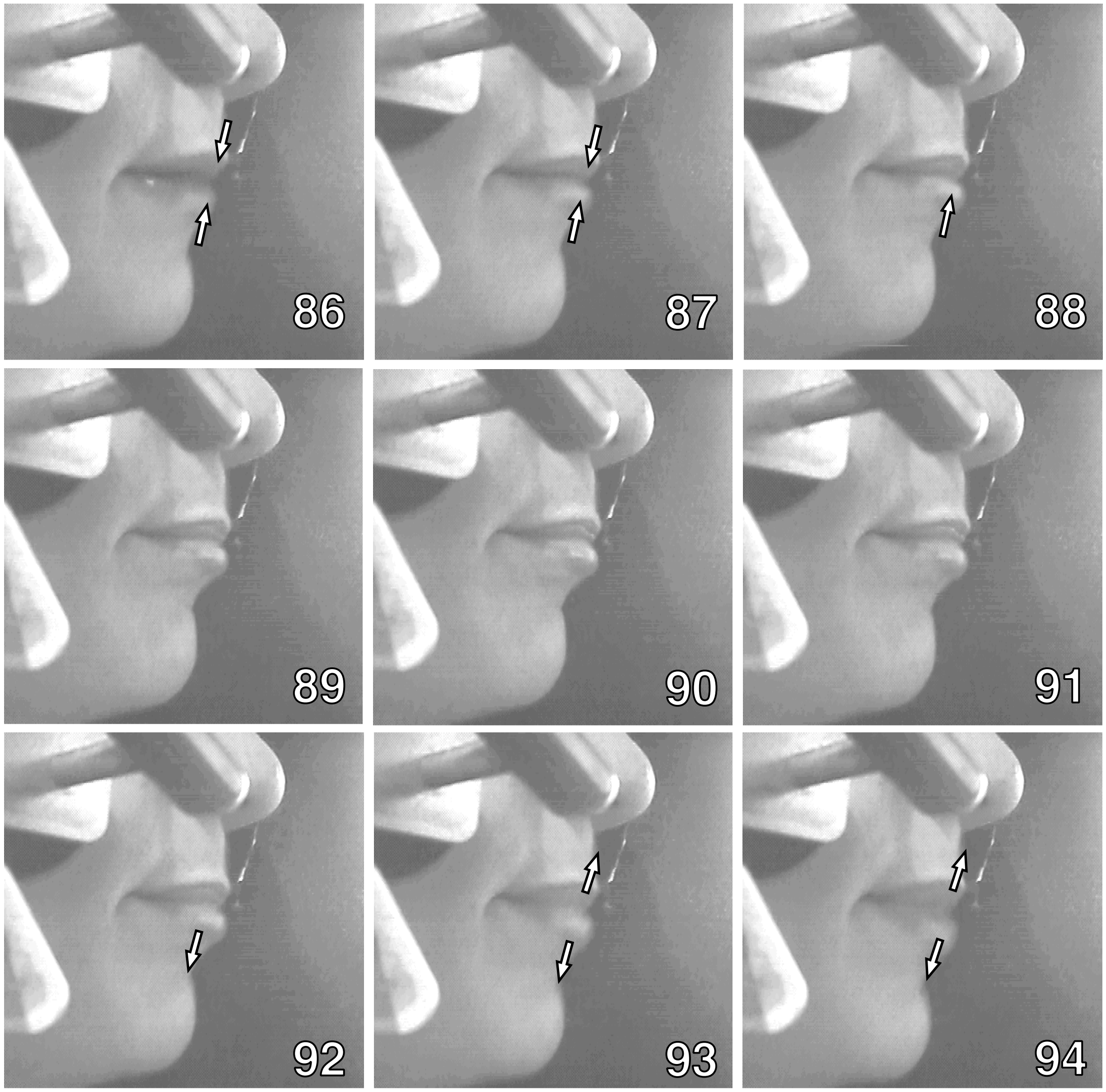


Figure 2.10: A series of nine camera frames during [p] from *είπα* [ˈi.**p**a] ‘I said…’ in the carrier phrase *Είπα* ______ *και πάλι*, as produced by speaker S10, showing the movement of the upper and lower lips toward closure (frames 86–87), during closure (frames 88–92), and out of closure (frames 93–94), with greater compression between the lips during closure occurring during frames 89–91.

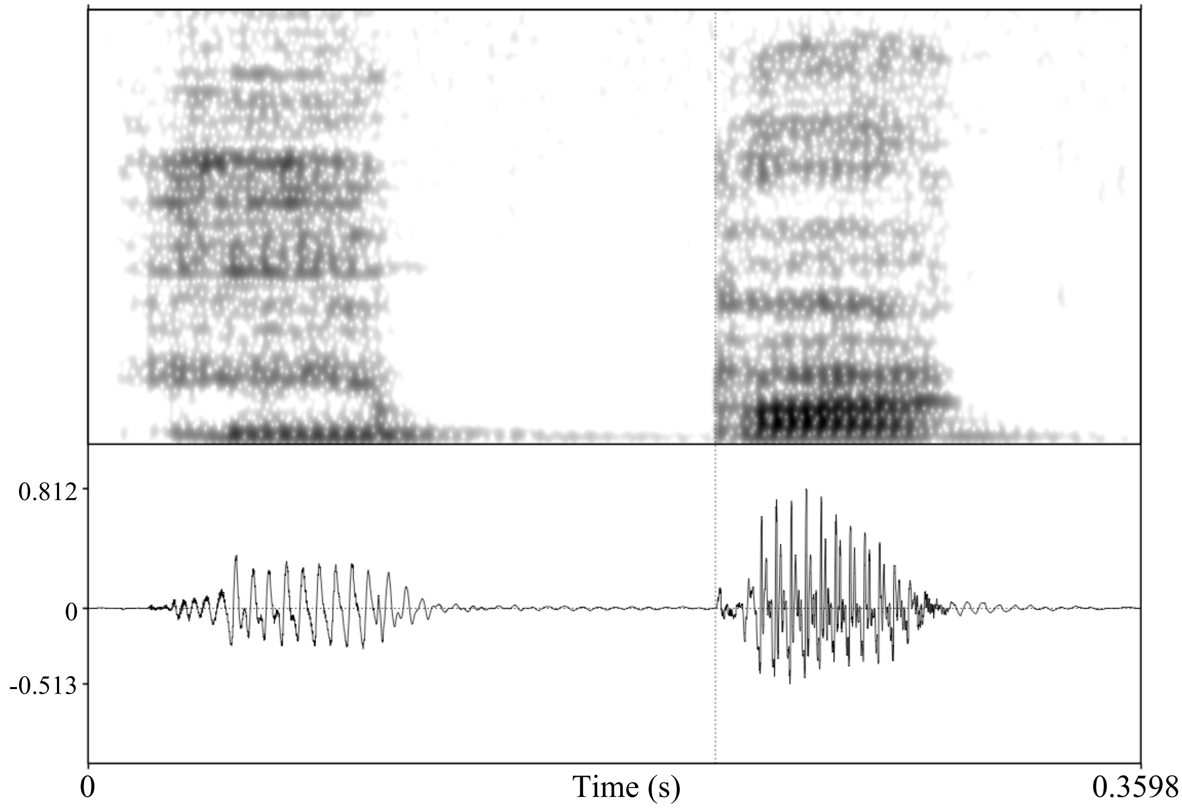


Figure 2.11: Identifying the acoustic correlate the release of the bilabial stop in *Είπα* [ˈi.**p**a] (sample taken from speaker S10). This is the same [p]-closure for which the camera images are shown in Figure 2.10. The vertical dotted line marks the onset of the acoustic release burst, which occurs at the same time as frame 93 in Figure 2.10.

For the analysis of the labial aperture appearing in lip-camera video, spatial points along the upper and lower lips were directly drawn onto the relevant video-frame images. For each speaker, labial aperture was measured by tracking the distance between these upper- and lower-lip points during labial constriction. The location of such points depended on whether the closure was a bilabial plosive, [p], or a labiodental fricative, [f]. For [p], the two closure points were defined as the frontmost point along the bottom edge of the upper lip and the frontmost point along the top edge of the lower lip, as seen from a profile view. This is shown in Figure 2.12. For [f], where only the movement of the lower lip was relevant to constriction achievement, aperture was measured as the vertical distance from the highest point of labial constriction along the front edge of the upper front incisors to the top edge of the lower lip. This is shown in Figure 2.13. A sample aperture plot for a single token of [pt] is presented in Figure 2.14.

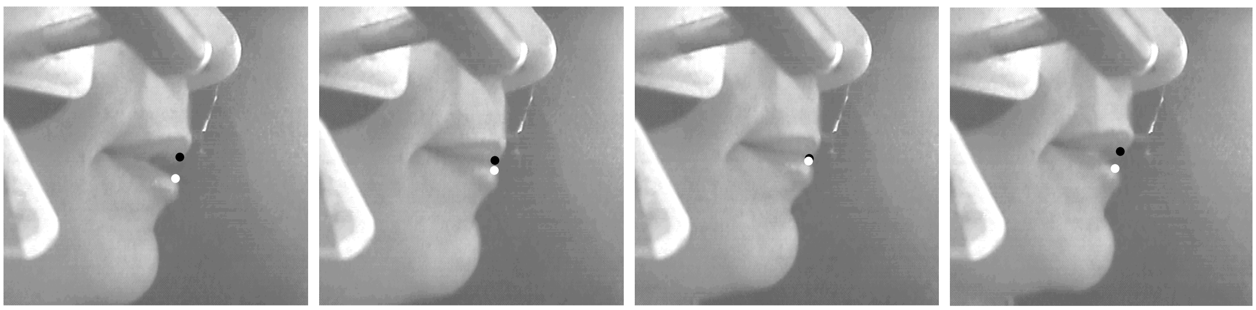


Figure 2.12: A series of frames during [p] in the word *πτήση* [ˈ**p**ti.si] ‘flight’, as produced by speaker S10. In order to demonstrate the entire movement, frames were taken 50-ms apart (20 fps). In each frame, the black dot indicates frontmost point along the lower edge of upper lip, while the white dot indicates the frontmost point along the upper edge of the lower lip. Aperture for labial plosive [p] is measured as the distance between these two points.

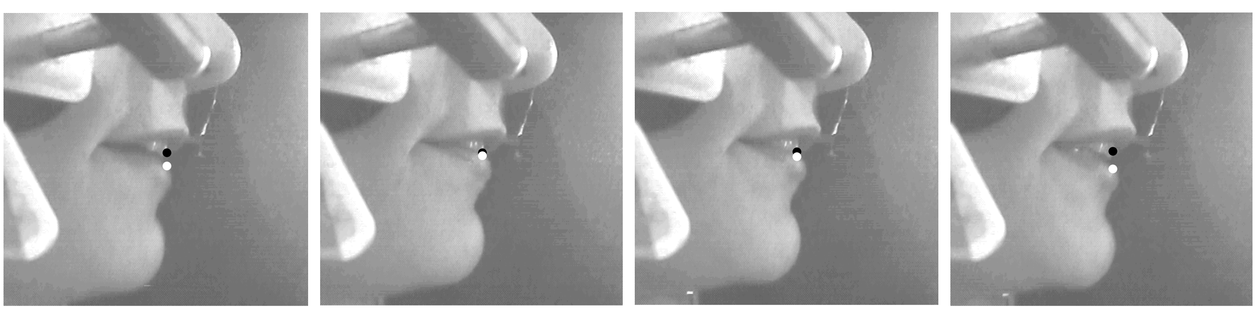


Figure 2.13: A series of frames during [f] in the word *φταίω* [ˈ**f**te.o] ‘(I) am at fault’, as produced by speaker S10 (frames were taken 50 ms apart = 20 fps). In each frame, the black dot indicates highest extent of lower-lip movement over the upper front incisors observed during labiovelar frication (this is the same location in all frames), and the white dot indicates the point along the upper edge of the lower lip located directly below each black dot. Aperture for labiovelars is measured as the distance between these two points.

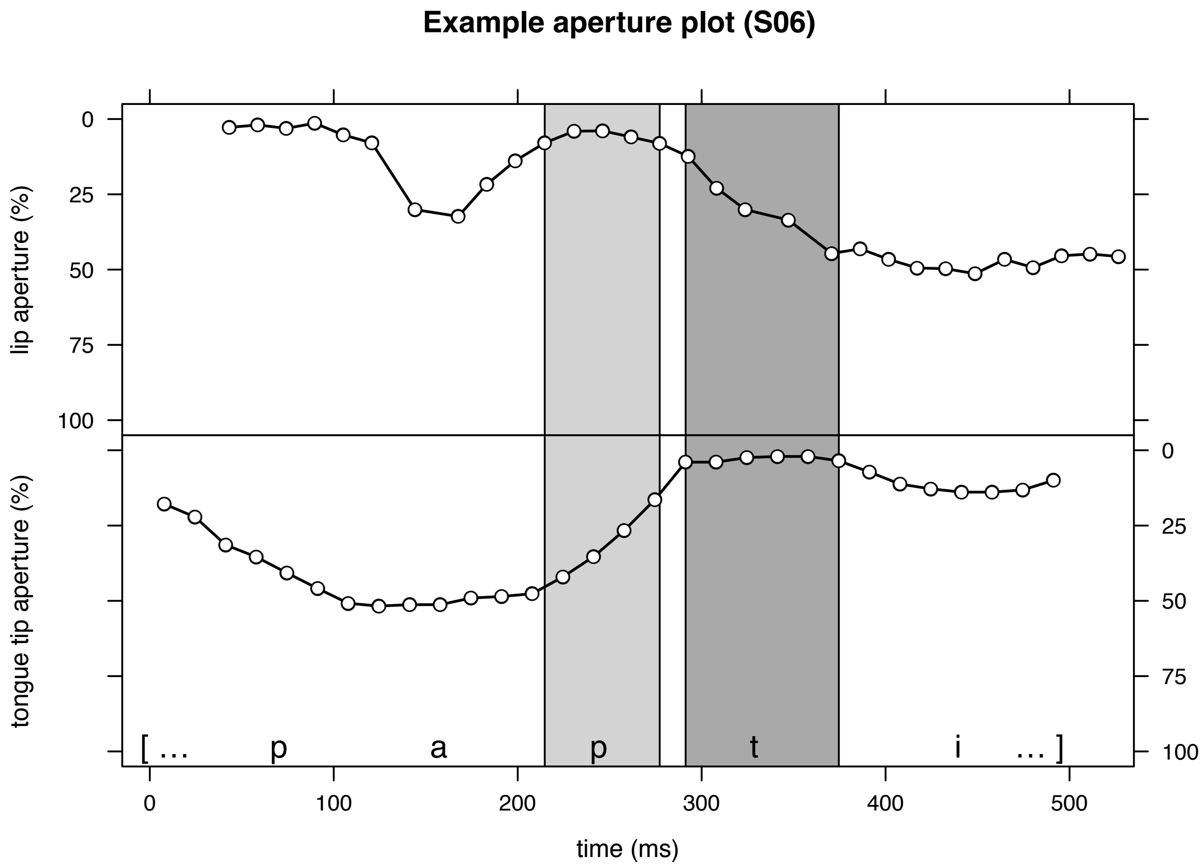


Figure 2.14: Lip (top) and tongue tip (bottom) aperture, measured in percent of total aperture range for each articulator, against time during the production of *(Εί)πα πτή(ση)…* [(ˈi).pa ˈ**pt**i.(si)…] by speaker S10. Vertical lines distinguish articulatorily defined boundaries for the closure intervals of the two segments [p] (light gray interval) and [t] (dark gray interval), discussed in Section §2.3.3).

2.3.3 Identifying gestural landmarks

For each CC production, gestural overlap between C1 and C2 was quantified using the trajectory graphs for the labial, coronal, and dorsal gestures. These trajectory graphs contained polynomial smoothing curves fitted to the raw percentage-aperture data at the tongue tip, tongue dorsum, and lips (upper lip and lower lip or lower lip to upper front teeth). Since the scan rate for both lingual and labial movement was fast (60 fps), applying these smoothing functions enabled an approximation of the velocity of aperture change over the course of an articulatory gesture. Figure 2.15 shows an example trajectory graph for TT and TD aperture in [kt]. Labial aperture measures were fit to smoothing curves in the same manner, as exemplified in Figure 2.16.

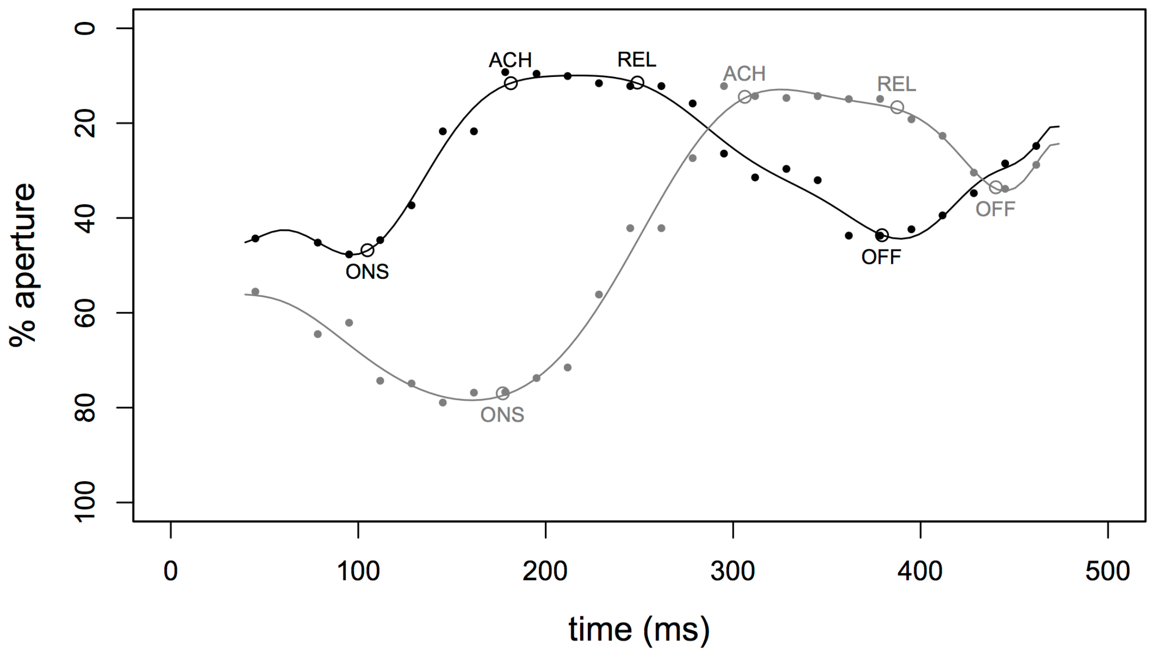


Figure 2.15: Measures for aperture over time at the tongue dorsum (black dots and curve) and tongue tip (gray dots and curve) during the sequence [kt] taken from a single token of the word *κτίσμα* [ˈ**kt**i.zma] 'building (n.)' produced by speaker S07. Slope-defined times for gestural onset, achievement, and release for C1 ([k]) and C2 ([t]) are indicated with empty circles and labeled accordingly.

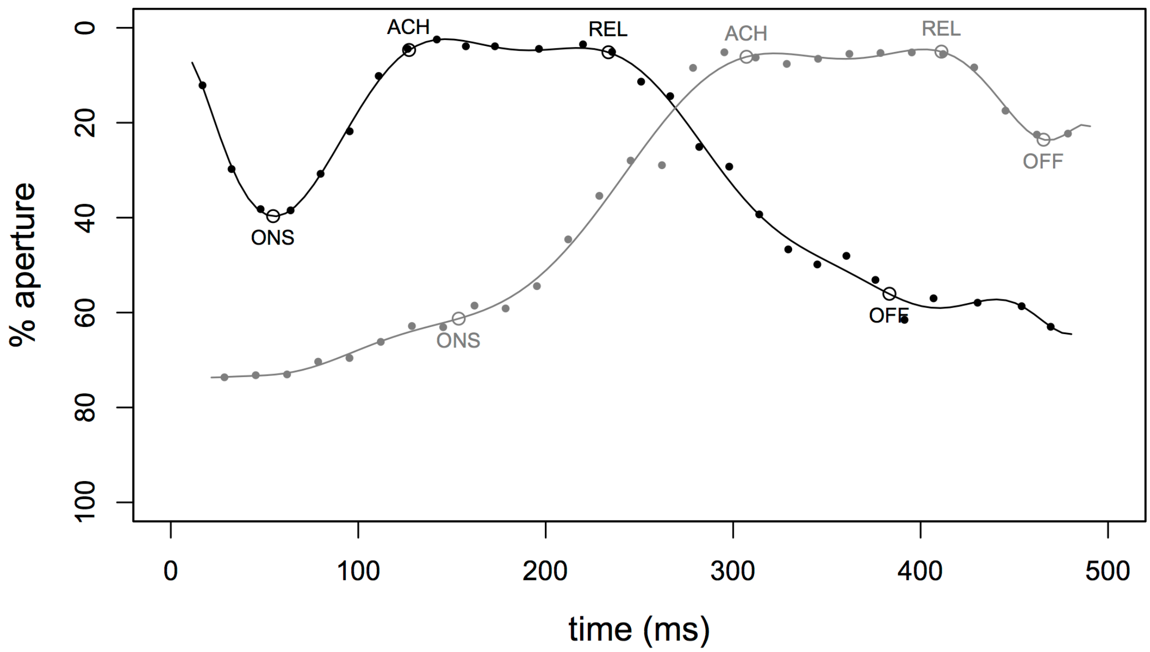


Figure 2.16. Measures for aperture over time at the lips (black dots and curve) and tongue tip (gray dots and curve) during the sequence [pt] taken from a single token of the word *πτήση* [ˈ**pt**i.si] 'flight' produced by speaker S07. Slope-defined times for gestural onset, achievement, and release for C1 ([p]) and C2 ([t]) are indicated with empty circles and labeled accordingly.

For the lingual or labial movements of C1 and C2 in each CC token, velocity values during interval of the gestural movement were measured by tracking the velocity of the smoothing curves in the corresponding trajectory graph. The maximum velocity for each gestural movement was thus based on the largest observed velocity magnitude for the movement. This maximum velocity value was used to identify gestural timepoints in a manner similar to that employed by Chitoran *et al.* (2002), in which gestural landmarks (onset, achievement, release, and offset) were defined as occurring at a threshold of 15–20% of maximum velocity of aperture change, depending on articulator (tongue tip, tongue dorsum, and lower lip: 15%; upper lip: 20%). For the analysis of data collected in this experiment, the velocity threshold percentage for all articulatory gestures in this study was the same (20%), in order to provide consistency across aperture measures that were all adjusted to a percent aperture scale regardless of the nature of the gesture. For all articulatory constrictions, gestural landmarks were identified in the following ways:

Onset time (ms): Time of onset for a given constriction was defined as the time at which percent aperture for the target articulator decreased at a trajectory-curve velocity that exceeded 20% of maximum velocity for that movement (that is, the time at which aperture increased at a rate that exceeded the 20% velocity threshold).

Achievement time (ms): Time of constriction achievement was defined as the time following constriction onset at which percent aperture decreased at a trajectory-curve velocity that fell below 20% of maximum velocity.

Release time (ms): Time of constriction release was defined as the time after achievement at which aperture increased at a trajectory-curve velocity that exceeded 20% of maximum velocity after the time of achievement.

Offset time (ms): Time of constriction offset was defined as the time following release at which aperture increased at a trajectory-curve velocity falling below 20% of maximum velocity at the end of gestural movement (that is, the time at which aperture decreased at a rate that fell below the 20% velocity threshold).

Example illustrations of how these gestural time points map onto trajectory graphs for CC sequences [kt] and [pt] are shown in Figures 2.15 and 2.16, respectively.

Additionally, a measure of lag between the C1 and C2 gestures aimed specifically at addressing CC-sequence perceptibility was calculated based on the critical time points identified within each CC token:

Intergestural lag (ms): Intergestural lag time was defined as the time of C2-constriction achievement minus the time of C1-constriction release. This measure specifies the temporal separation between the intervals of maximum constriction for C1 and C2 and thus reveals the degree of intergestural overlap that might, for example, contribute to acoustic masking. When lag measures are negative, the achievement intervals for C1 and C2 overlap temporally, potentially resulting in acoustic masking during the overlap interval.

2.4 Predictions

As explained in Chapter I, the perceptual-recoverability and biomechanical hypotheses make different predictions for the effects of C1 and C2 manner on gestural overlap and the interaction of these effects with the influence of place order on gestural overlap. In general, under perceptual recoverability, the timing of gestures depends on whether gestural overlap could lead to acoustic masking; that is, when masking due to overlap is likely, speakers should compensate by with increased gestural separation or intergestural lag. Under a biomechanical account, gestural timing is governed by the physical capabilities of the articulators, especially when gestural movements for C1 and C2 are interdependent. When the motor demands for each of the two gestures conflict, speakers are expected to produce C1 and C2 in sequence rather than simultaneously, resulting in longer intergestural lag.

The following sections lay out the separate sets of predictions for initial Greek CC clusters made by the two theoretical approaches to gestural coordination. These predictions are illustrated as schematic aperture plots in Figure 2.17.

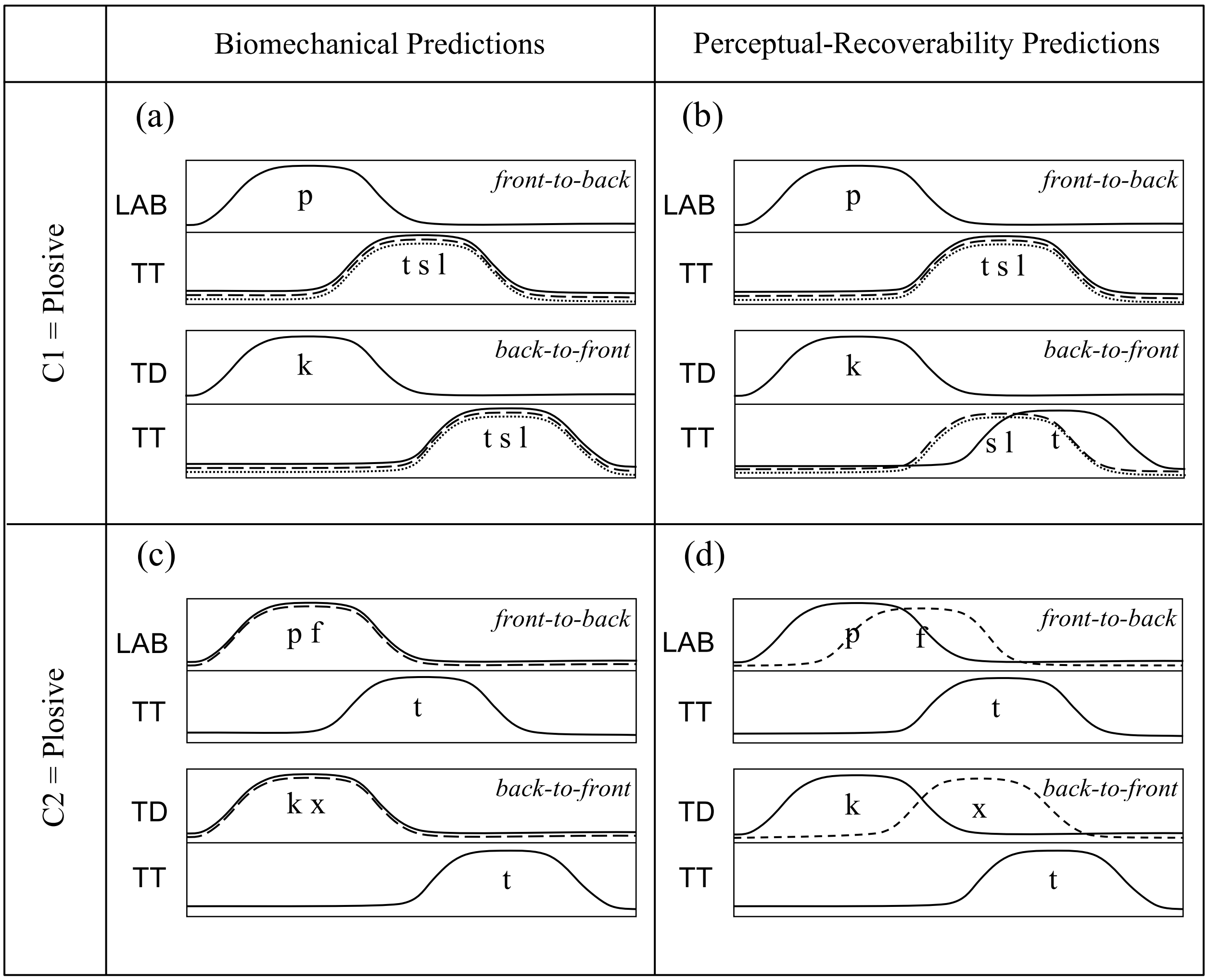


Figure 2.17. Schematic diagrams of the biomechanical (a,c) and perceptual-recoverability (b,d) predictions for the coordination of word-initial Greek CC sequences. In each diagram, the vertical axis represents aperture (0% on top) and the horizontal axis represents time. Graphs (a) and (b) show comparisons between plosive-plosive (solid curve), plosive-fricative (dashed), and plosive-lateral (dotted) productions in front-to-back vs. back-to-front sequences, where "LAB", "TT", and "TD" refer the lips, tongue tip, and tongue dorsum, respectively. Graphs (c,d) show comparisons between plosive-plosive (solid) and fricative-plosive (dashed).

Due to difficulties in the identification of the gestural landmarks corresponding to the Greek rhotic tap [ɾ], the predictions and results for plosive-tap sequences [pɾ kɾ] are omitted from this dissertation. Instead of the term "plosive-liquid", I henceforth use the phrase "plosive-lateral" to denote CC sequences [pl kl].

2.4.1 Perceptual-recoverability predictions

As explained in Section §1.6, the likelihood of perceptual masking should be the greatest when the manners of C1 and C2 are both plosive and when the place order is back-to-front. When overlapped with a following plosive constriction for C2, a plosive constriction for C1 may be released with cues that are potentially obscured by airstream blockage created by C2 closure. However, place of articulation also matters here: for plosive-plosive sequences, the second closure should primarily mask release of the first of the sequence when the location of the second closure is in front of the first closure. If instead the second closure occurs behind the first closure, then the first closure is generally assumed to be released audibly during a simultaneous posterior closure. Thus, for Greek plosive-plosive sequences, the perceptual-recoverability hypothesis predicts that speakers will produce greater lag in back-to-front [kt] than in front-to-back [pt] to avoid potential masking of the plosive-C1 release in [kt] (Figure 2.17b).

If C1 is a plosive and C2 is a fricative, the overall likelihood of masking decreases, compared to plosive-plosive sequences. In the case of front-to-back sequences, less acoustic masking resulting from C1-C2 overlap is expected when C2 is a fricative ([ps]) than when it is a plosive ([pt]), since the acoustic cues to the release of labial closure for [p], i.e., its release burst and aspiration, are less obscured by the partial constriction gesture for [s]. For back-to-front sequences, cues to a velar release for [k] should also be present in the acoustic stream despite overlap with a simultaneous fricative [s] constriction. That is, a constriction in the coronal region for the fricative should not mask the acoustic signature of the release of a velar-plosive closure. Thus, in both place orders of a plosive-fricative combination, a delay between the constriction intervals of C1 (plosive) and C2 (fricative) is not needed to prevent full acoustic masking of the C1 gesture. However, it is important to consider whether any simultaneous fricative constriction could alter the salience of the release of the C1 plosive. If this is the case, speakers who seek to perceptually enhance their productions might be expected to respond by increasing the intergestural lag interval so that the fricative constriction does not substantially reduce the perceptual salience of the C1 release burst.

The behavior of plosive-lateral sequences should presumably resemble that of plosive-fricative sequences. For Greek plosive-lateral sequences [pl kl], constriction for C2 in the coronal region during the release of C1 are not expected to substantially obscure C1's release cues, regardless of place order, since lateral production does not involve a complete obstruction of the vocal tract. However, as with plosive-fricative sequences, if for some reason the acoustic salience of C1 release is reduced by a concurrent lateral constriction for C2, then perceptual recoverability predicts timing adjustments in the form of increased lag between C1 and C2.

Given these predictions for the effect of C2 manner, the perceptual recoverability hypothesis predicts an interaction between the effects of place order and C2 manner, such that Greek speakers will produce greater intergestural lag in plosive-plosive [kt] than in plosive-fricative [ks] and plosive-lateral [kl], but similar durations of lag in plosive-plosive [pt], plosive-fricative [ps], and plosive-lateral [pl] (Figure 2.17b).

As discussed in Section §1.5, during fricative-plosive sequences [ft xt] (Figure 2.17d), achievement of C2 constriction during the interval of C1 constriction will not completely mask acoustic cues (frication) to C1, provided that the constriction intervals for C1 and C2 are not initiated simultaneously. That is, if listeners hear sufficient frication to accurately perceive fricative C1, then partial (but not complete) overlap between the constriction intervals of C1 and C2 should render C1 perceptible. For this reason, there is expected to be less pressure for speakers to avoid constriction overlap in fricative-plosive than plosive-plosive sequences, and intergestural lag between C1 and C2 should be shorter in fricative-plosive sequences [ft xt] than in plosive-plosive sequences [pt kt]. This effect should be the same in both place orders.

While not pursued in the current study, an examination of the acoustic data corresponding to the CC-sequence productions is needed to test the perceptual-recoverability predictions made here. If, in their productions, speakers adjust the degree of intergestural overlap for perceptual reasons, then their productions should consequently contain sufficient acoustic cues to both overlapped and non-overlapped CC sequences. Additionally, appropriate perceptual testing is needed to assess the extent to which speakers' (arguably) perceptually-oriented adjustments to gestural timing actually improve listeners' recovery of phonetic information.

2.4.2 Biomechanical predictions

Like the perceptual-recoverability hypothesis, the biomechanical perspective on gestural coordination potentially predicts an effect of place order on C1-C2 overlap, i.e., less intergestural lag in labial-coronal ("front-to-back") sequences [pt ps pl ft] than in dorsal-coronal ("back-to-front") sequences [kt ks kl xt]. In this case, though, the order effect is due to the likelihood that in dorsal-coronal sequences, the C1 and C2 gestures will conflict with each other and consequently cannot substantially overlap, as discussed in Section §1.4. However, because, under a strict biomechanical account, acoustic masking effects are not expected to influence the degree of lag between the consonantal gestures, there is no predicted interaction between place order and manner. That is, there is no obvious, inherent biomechanical reason for dorsal-coronal [kt ks kl xt] sequences to exhibit more of an effect of manner than for labial-coronal [pt ps pl ft] sequences to do so (Figures 2.17a and 2.17c). Alternatively, if, in dorsal-coronal [kt ks kl xt] sequences, the production of constriction gesture for C2 depends on its manner—e.g., if the movement of the tongue body from [k] to [s] or from [k] to [l] is shorter or faster than the movement of the tongue body from [k] to [t]—then an interaction between place order and C2 manner is expected. Without previous findings in the literature on CC sequences that speak to this potential outcome, the biomechanical predictions here tentatively assume no substantial coordinative differences between C1 and plosive versus fricative versus lateral gestures for C2 (but see Section §4.2.2). Consequently, the largest expected differences in intergestural lag should therefore occur as a function of place order or, more precisely for the Greek CC sequences, articulator independence-interdependence.

The predictions following from the biomechanical and perception-recoverability accounts of gestural coordination from Figure 2.17 and discussion in this section are summarized in Table 2.2.

		Biomechanical predictions		*Perceptual-recoverability predictions*	
Word-initial position	**C1 = plosive**	• Greater lag in dorsal-coroanl (“back-to-front”) than in labial-coronal (“front-to-back”).	[kt] > [pt] [ks] > [ps] [kl] > [pl]	• **Plosive C2**: Greater lag in back-to-front than in front-to-back. **Fricative, lateral C2**: Equal lag regardless of place order.	[kt] > [pt] [ks] = [ps] [kl] = [pl]
		• Equal lag regardless of C2 manner. (tentative)	[kt] = [ks] = [kl] [pt] = [ps] = [pl]	• **Back-to-front**: Greater lag when C2 is plosive than fricative or lateral. **Front-to-back**: Equal lag regardless of C2 manner.	[kt] > [ks kl] [pt] = [ps] = [pl]
	C2 = plosive	• Greater lag in dorsal-coronal (“back-to-front”) than in labial-coronal (“front-to-back”).	[kt] > [pt] [xt] > [ft]	• **Plosive C1**: Greater lag in back-to-front than in front-to-back.	[kt] > [pt]
		• Same manner pattern in dorsal-coronal (“back-to-front”) as labial-coronal (“front-to-back”).	[kt] = [xt] [pt] = [ft]	• **Back-to-front**: Greater lag for plosive C1 than for fricative C1. **Front-to-back**: Greater lag for plosive C1 than for fricative C1.	[kt] > [xt] [pt] > [ft]

Table 2.2: Summary of predicted effects for the biomechanics and perceptual-recoverability predictions for the coordination of Modern Greek CC sequences in word-initial contexts.

CHAPTER III
Results

3.1 Speaker Exclusion

Table 3.1 lists the ten recorded speakers, their sex and age, and a short statement of their language backgrounds. Most recruited speakers were male, although male speakers were not specifically targeted in this study. Because the speakers were recruited at the University of Michigan (Ann Arbor, MI), all participants, although native speakers of Greek, also spoke English. These native Greek-speaking participants fell into two categories: Greek-dominant speakers who grew up in Greece and learned English later in life, and early bilinguals in both Greek and English who grew up in southeastern Michigan and belonged to the relatively large Greek-speaking community in the Detroit, MI area. The only participant from the latter group ultimately used in the study is speaker S08, who grew up speaking Standard Greek in the home, lived in Greece for two years, and travels back to Greece each summer for work.

Of the ten speakers recruited for the study, two were excluded from the final analysis due to issues with the quality of ultrasound data collected from those speakers. Data from speaker S04 were excluded because in many contexts the tongue tip (TT) was outside of the viewable area of the ultrasound image. Data from speaker S09 were excluded because, in recording sets containing items with dorsal gestures, the tongue dorsum (TD) disappeared from the viewable area in the ultrasound image as it raised to a dorsal-constriction position.

Speaker	Sex	Age	Language background (hometown)	Data used in the study?
S01	male	29	Greek-dominant (Athens, Greece)	YES
S02	male	20	Greek-dominant (Athens, Greece)	YES
S03	male	26	Greek-dominant (Athens, Greece)	YES
S04	male	31	Greek-dominant (Cyprus)	NO: TT outside of viewable area
S05	male	37	Greek-dominant (Athens, Greece)	YES
S06	male	26	Greek-dominant (Gargaliánoi, Greece)	YES
S07	female	26	Greek-dominant (Athens, Greece)	YES
S08	female	26	Greek-English bilingual (Detroit, MI)	YES
S09	male	20	Greek-English bilingual (Detroit, MI)	NO: max. extent of TD not visible
S10	female	34	Greek-dominant (Athens, Greece)	YES

Table 3.1: Summary of participants in the experiment.

3.2 Statistical procedures

Due to the wide range of intergestural lag durations produced by the different speakers, testing for lag effects across speakers required conversion of each speaker's lag times into normalized, *z*-score values. To determine whether intergestural lag (as a *z*-score) was influenced by place of articulation order and of C1 and C2 manner, I computed three types of linear mixed-effects models (LMMs) using the lmer() function in the R package "lme4" (Bates *et al.*, 2011). These model types were: 1) a simple model with the fixed effect *place order* (front-to-back, back-to-front), 2) a model with fixed effects *place order* and *C1 manner* (plosive, fricative) and the interaction between these two effects, and 3) a model with fixed effects *place order* and *C2 manner* (plosive, fricative, lateral) and the interaction between these two effects. *Speaker* and *item* (individual words) were included in each model as random intercepts. Rather than reporting the results of an omnibus test, LMMs show all main effects as paired *t*-test comparisons. For each model, *p*-values indicating the level of significance for each comparison were then estimated using Markov chain Monte Carlo (MCMC) simulations with the pvals.fnc() function from the R package "languageR" (Baayen, 2008).

Additionally, to test for effects within the data for individual speakers, I computed the same three types of LMMs on separate sets of intergestural lag data from each speaker, with only a single random effect of *item* and the corresponding fixed effects for each model type: 1) *place order* only, 2) *place order* and *C1 manner* and their interaction, and 3) *place order* and *C2 manner* and their interaction. These LMMs were computed in the same manner as the three LMMs performed across data for all speakers. Because of the lack of comparisons between the productions of different speakers, intergestural lag times were not converted into *z*-scores for the individual-speaker analyses.

3.3 Order of Place of Articulation

Recall from Section §2.4 that both the biomechanical and the perceptual-recoverability hypotheses, as interpreted here for labial-coronal and dorsal-coronal sequences, predict the same outcome for the effect of order of place of articulation (front-to-back, back-to-front) on the intergestural lag between consonants in plosive-plosive CC sequences, but not for some sequences in which one of the consonants is non-plosive:

Prediction 1b (Biomechanics): Intergestural lag, i.e., lag between the constriction intervals of C1 and C2, should be longer in back-to-front, dorsal-coronal [kt] than in front-to-back, labial-coronal [pt], due to the different (active) articulators involved for each of the two place orders. For the same reason, in broad terms, the place-order effect should also hold across the other manners ([ks] > [ps], [kl] > [pl], and [xt] > [ft]).

Prediction 1p (Perceptual Recoverability): Intergestural lag in back-to-front [kt] should be longer than that in front-to-back [pt] due to differences in the likelihood of masking when the constriction intervals of C1 and C2 overlap. However, this place-order pattern should not hold (or should not hold to the same extent) for non-plosive–plosive sequences [ft xt] or plosive–non-plosive sequences [ps ks pl kl], for which masking due to C1 and C2 overlap is unlikely.

Longer intergestural lags for back-to-front, dorsal-coronal sequences than for corresponding front-to-back, labial-coronal sequences were upheld across all speakers as shown by the results of the linear mixed-effects model for place order, given in Table 3.2. The effects of place order are graphically represented in Figure. 3.1. Thus, the prediction for the main effect of place order was upheld.

Comparison			Estimates (β)		t	p
Place Order	Plosive-plosive	[pt]~[kt]	[pt]: -0.29 (0.06)	[kt]: 0.79 (0.07)	14.86	0.0001
	Plosive-fricative	[ps]~[ks]	[ps]: -0.51 (0.07)	[ks]: 0.39 (0.10)	8.62	0.0001
	Plosive-lateral	[pl]~[kl]	[pl]: -0.47 (0.14)	[kl]: 0.64 (0.14)	8.13	0.0001
	Fricative-plosive	[ft]~[xt]	[ft]: -0.72 (0.09)	[xt]: 0.24 (0.12)	7.65	0.0001

Table 3.2: Paired comparisons testing Predictions 1b and 1p from the linear mixed models fit to the *z*-scores for intergestural lag durations, pooled across speakers, in plosive-plosive [pt kt], plosive-fricative [ps ks], plosive-lateral [pl kl], and fricative-plosive [ft xt] sequences. Standard errors for each estimate are reported in parentheses.

As discussed in Sections §1.4 and §2.4, the biomechanical and the perceptual-recoverability hypotheses are both consistent with greater back-to-front (dorsal-coronal) than front-to-back (labial-coronal) intergestural lag in plosive-plosive sequences. According to the biomechanical hypothesis, gestural overlap is more pervasive in labial-to-coronal [pt] than in dorsal-to-coronal [kt] because in [pt] the gestures are made by separate articulators, whereas in

[kt] the gestures are made by the same articulator. By contrast, the perceptual-recoverability hypothesis asserts that the difference in lag is due to speakers' effort to avoid masking in back-to-front contexts, but not in front-to-back contexts.

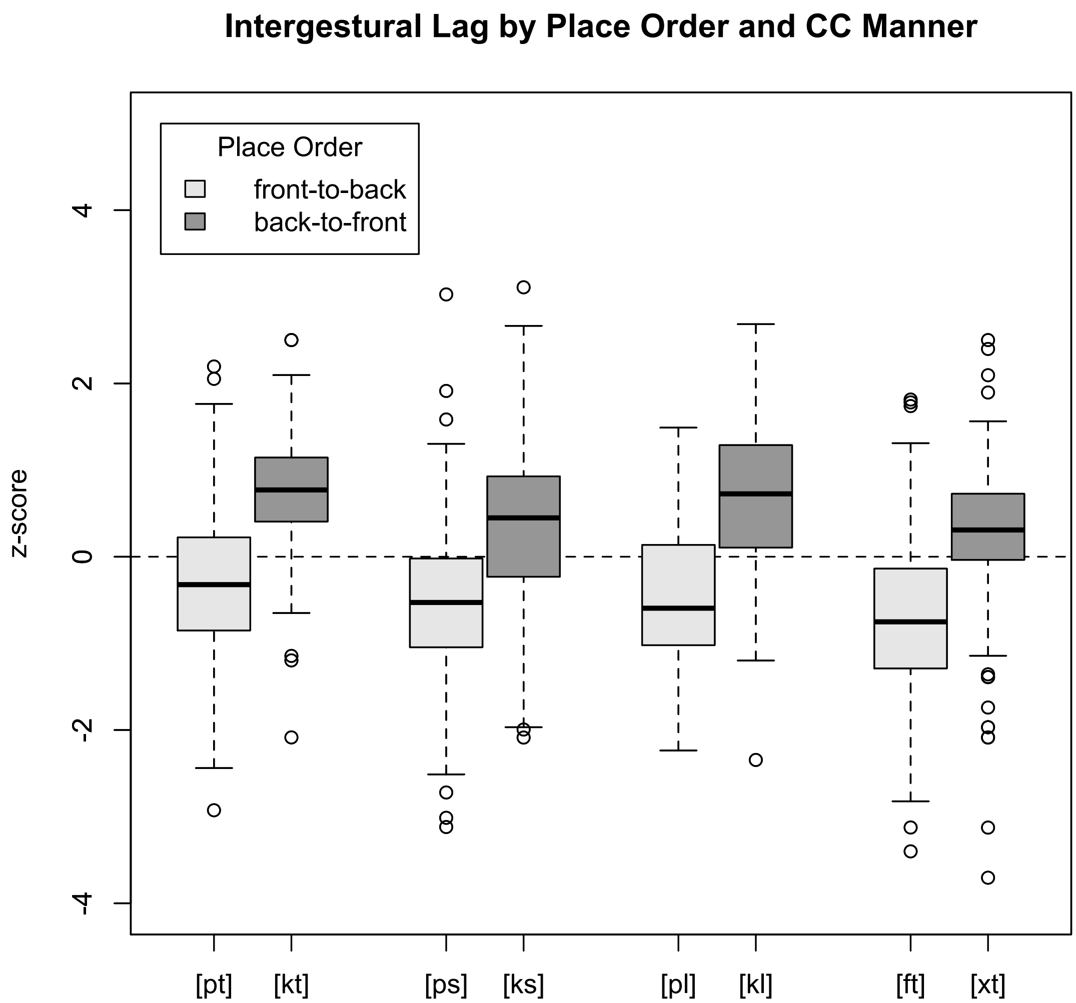


Figure 3.1: *z*-scores of intergestural lag for for CC sequences [pt kt ps ks pl kl ft xt] across all speakers, by place order.

The LMM estimates for intergestural lag in front-to-back (labial-coronal) sequences [ps pl ft] and their back-to-front (dorsal-coronal) counterparts [ks kl xt] have the same pattern that was observed for the plosive-plosive [pt kt] sequences. These results are consistent with Chitoran and Goldstein's (2006) finding that the place-order effect is generalizable to manner sequences other than plosive-plosive—in their study, to plosive-liquid and liquid-plosive sequences. These findings across manner combinations are also consistent with the biomechanical approach, which identifies the interdependence of the C1 and C2 articulators as a primary cause of longer lag in dorsal-coronal than labial-coronal sequences. In terms of the perceptual-recoverability approach, I have speculated that, although it is less clear whether the place-order effect should extend to manners other than plosive-plosive, we should minimally not expect as strong of an effect in these sequences as in plosive-plosive [pt kt].

According to the LMM for the place-order effect, the random effect of *speaker* contributed 36.1% of the overall variance in the intergestural lag data for all CC sequences, whereas the effect of *item* contributed 14.4% of the overall variance. To evaluate the variation between speakers, LMMs for *place order* were separately performed on the data from each of the eight speakers. Full analyses per speaker are provided in Sections §B.1–B.4 in the appendix, and the bar plots comparing place order by speaker, for each manner type, are presented in Figures 3.2–3.5. In this study, although the main effect of place order was found for all of the CC sequences, individual speakers' LMM results were more variable. However, as shown in Figure 3.2 and in the Appendix, §B.1, all speakers showed the group pattern for the plosive-plosive sequences.

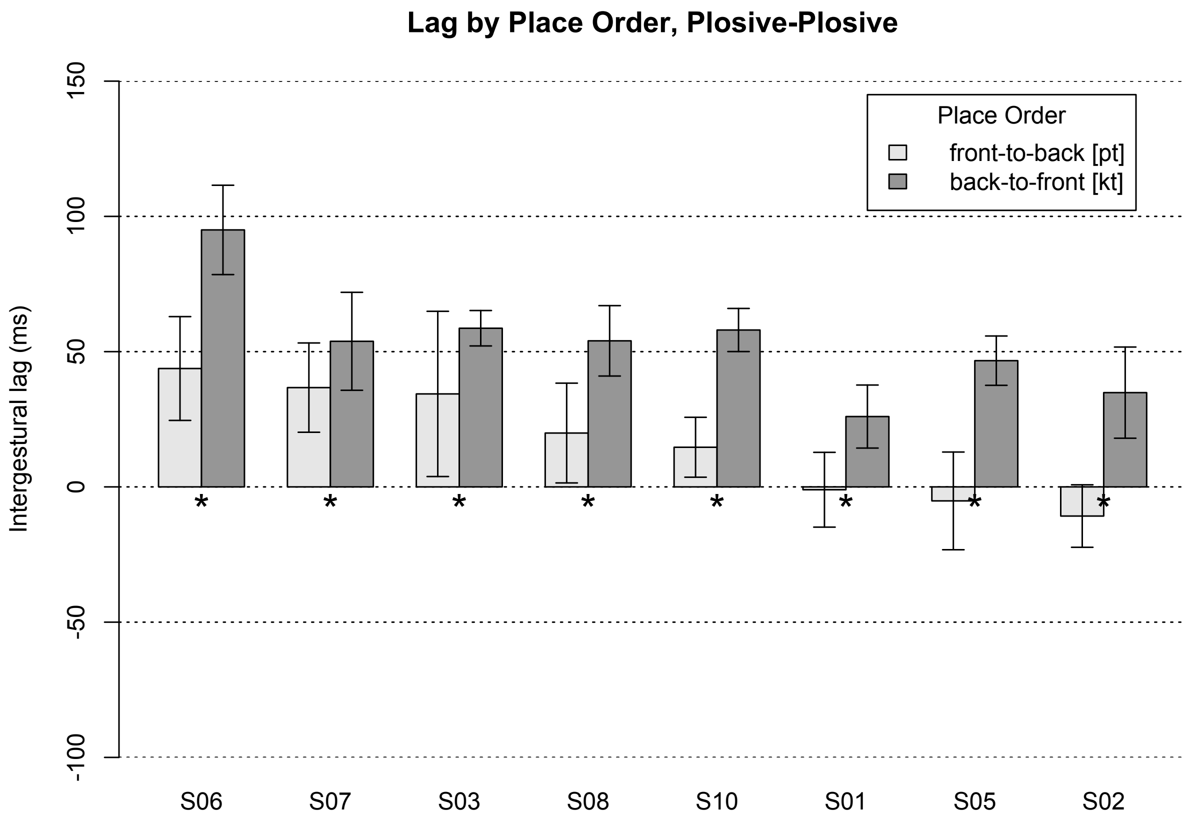


Figure 3.2: Intergestural lag durations (ms) for plosive-plosive sequences [pt kt], by place order and speaker. Speakers are ordered from longest to shortest mean lag duration in [pt] sequences. Asterisks signify a significant effect of place order, at $p<0.05$. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

Although most speakers' data for the plosive–non-plosive and non-plosive–plosive sequences conformed to the plosive-plosive lag pattern, some speakers' productions did not show this pattern to the 0.05 level of statistical significance. In plosive-fricative sequences, presented in Figure 3.3 and in the Appendix, §B.2, all speakers except for S03 and S07 produced significantly longer intergestural lag in back-to-front [ks] than front-to-back [ps].

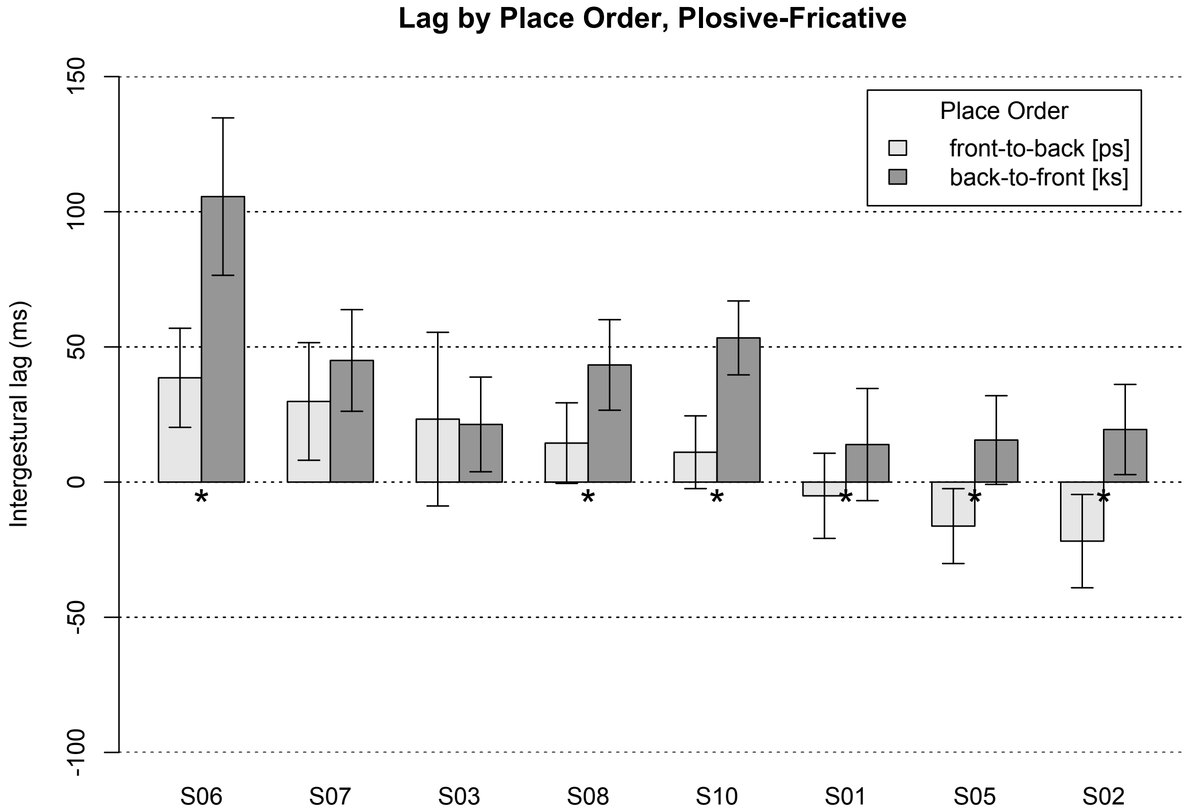


Figure 3.3: Intergestural lag durations (ms) for plosive-fricative sequences [ps ks], by place order and speaker. For consistency with Figure 3.2, speakers are ordered from longest to shortest mean lag duration in [pt] sequences. Asterisks signify a significant effect of place order, at $p<0.05$. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

For plosive-lateral contexts, presented in Figure 3.4 and in the Appendix, §B.3, six of the eight speakers (S01, S02, S05–S07, S10) produced significantly longer lag in back-to-front [kl] than in front-to-back [pl]. While speakers S03 and S08 also appear to have this pattern in their lag durations, the differences in their lag estimates were not significant (S03: 48.1 ms in [pl], 61.1 ms in [kl], $p = 0.404$; S08: 7.7 ms in [pl], 29.2 ms in [kl], $p = 0.110$).

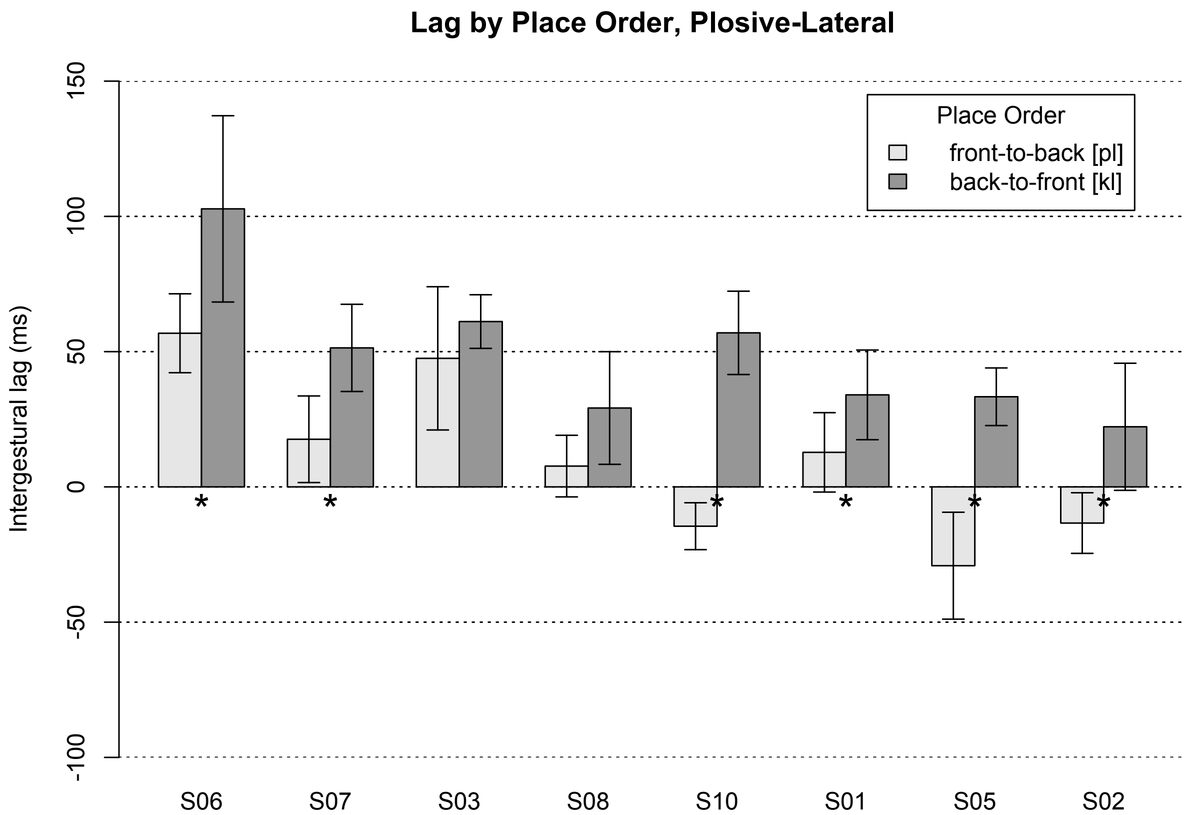


Figure 3.4: Intergestural lag durations (ms) for plosive-lateral sequences [pl kl], by place order and speaker. For consistency with Figure 3.2, speakers are ordered from longest to shortest mean lag duration in [pt] sequences. Asterisks signify a significant effect of place order, at $p<0.05$. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

For fricative-plosive contexts, shown in Figure 3.5 and in Appendix §B.4, only half of the speakers (S02, S05, S07, S10) produced significantly longer lag in back-to-front [xt] than in front-to-back [ft], although once again the speakers who did not show a significant difference in the model had intergestural lag estimates that followed the overall group pattern (S01: -8.6 ms in [ft], 6.6 ms in [xt], $p = 0.448$; S03: 13.3 ms in [ft], 38.7 ms in [xt], $p = 0.122$; S06: 56.8 ms in [ft], 68.3 ms in [xt], $p = 0.083$; S08: 29.5 ms in [ft], 37.3 ms in [xt], $p = 0.2560$).

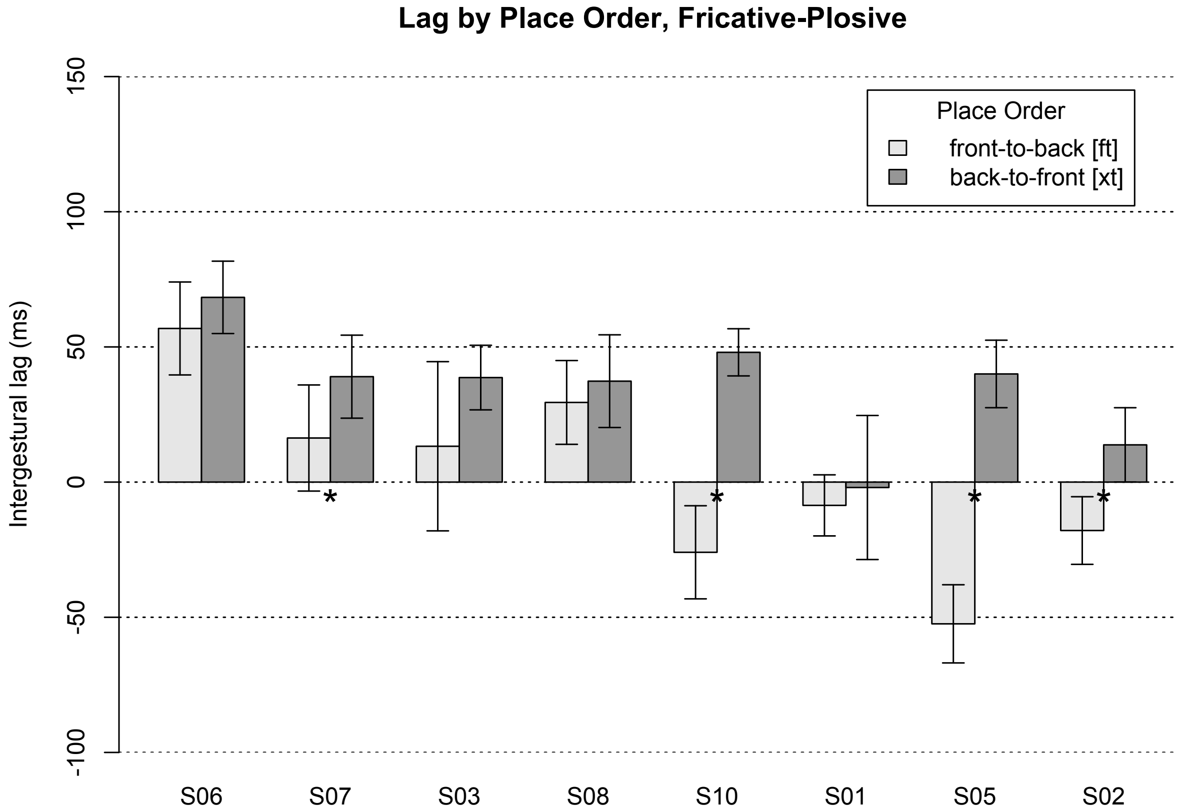


Figure 3.5: Intergestural lag durations (ms) for fricative-plosive sequences [ft xt], by place order and speaker. For consistency with Figure 3.2, speakers are ordered from longest to shortest mean lag duration in [pt] sequences. Asterisks signify a significant effect of place order, at $p<0.05$. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

Summary: The data for all speakers in this study reveal that, consistent with theoretical predictions, measures for lag between the constriction intervals of C1 and C2 in plosive-plosive sequences were longer in back-to-front [kt] than in front-to-back [pt]. The same effect of the order of place of articulation was observed across plosive-fricative [ps ks], plosive-lateral [pl kl], and fricative-plosive [ft xt] clusters, although the differences between place orders did not reach significance in each individual speaker's LMMs for these manner types. These results also show that the magnitude of the place order effect is relatively large, with an average difference in intergestural lag of 34.9 ms between the two place orders, across all manner types ([kt] − [pt]: 36.8 ms; [ks] − [ps]: 30.3 ms; [kl] – [pl]: 38.3 ms; [ft] − [xt]: 34.0 ms).

3.4 Relation between Place Order and C1/C2 Manner

Recall, from Section §2.4, the following predictions about the relation between place order and C1/C2 manner:

Prediction 2b (Biomechanics): If speakers' productions of CC sequences are timed primarily according to the independence of the articulators involved, then there should be no consistent influence of manner on intergestural lag in either labial-coronal [pt ps pl ft] or dorsal-coronal [kt ks kl xt] sequences. However, if manner-dependent articulatory constraints apply to lingual-lingual sequences, such as greater dorsal-to-coronal excursion for [kt] than for [ks], then a corresponding C2 manner effect on lag is expected only for these sequences.

Prediction 2p (Perceptual Recoverability) If speakers' productions are influenced by perceptual-recoverability factors, then intergestural lag across manners should differ in the two place orders, namely: for front-to-back sequences, fricative-plosive [ft] should have shorter intergestural lag than that of all other manners [pt ps pl], whereas for back-to-front sequences, plosive-plosive [kt] should have longer intergestural lag than that of all other manners [ks kl xt].

For the perceptual-recoverability hypothesis, the emergence of C1- and C2-manner effects on intergestural lag should depend at least to some extent on whether individual speakers produce positive or negative lag durations. Given that the duration of intergestural lag was measured as the interval from the release of C1 constriction to the achievement of C2 constriction, positive lag means that C1 was released before C2 was achieved, whereas negative lag means that C2 was achieved before C1 was released. Because substantial acoustic masking of the sort described in Sections §1.4.1 and §2.4.1 is expected to occur when the constriction intervals of the C1 and C2 gestures overlap, the perceptual-recoverability hypothesis makes the above predictions for the effects of place and manner only when (or at least most clearly when) speakers produce some degree of overlap between C1 and C2 constrictions, that is, negative, or near-negative, intergestural lag values.

To assess the effects of place order and C1 and C2 manner, two LMMs—one testing place and C1 manner and the other place and C2 manner—were performed, with *speaker* and *item* as random factors. For each LMM, the interaction between place and manner effects was also computed. The results of these models are shown in Table 3.3; contrasts in place order and C1 or C2 manner, across all speakers, are shown in Figure 3.6. To obtain the results for all possible pairwise comparisons in the LMM involving a factor for C2 manner, the reference factor was changed to [pt] for the comparisons [pt]~[ps] and [pt]~[pl], to [ps] for the [ps]~[pl], to [kt] for [kt]~[ks] and [kt]~[kl], and [ks] for [ks]~[kl].

The analysis for the fixed effects of *place order* and *C1 manner* across plosive-plosive [pt kt] and fricative-plosive [ft xt] sequences shows that, in the front-to-back context, plosive-plosive [pt] was produced with significantly larger *z*-score values for lag than fricative-plosive [ft], with estimates of -0.293 for [pt] and -0.717 for [ft]. In the back-to-front context, the difference between C1 manners were in the same direction, with estimates of 0.797 for plosive-plosive [kt] and 0.243 for fricative-plosive [xt]. Consequently, there was not a significant interaction between *place order* and *C1 manner*. Although the *z*-score estimates presented in Table 3.3 do not indicate whether intergestural lag was negative, i.e., whether the constriction intervals for C1 and C2 overlapped, mean lag values for the dorsal-coronal [kt xt] sequences produced by individual speakers (Figure 3.9) were predominantly positive, indicating little to no overlap (as defined here) between the C1 and C2 constrictions.

	Comparison		Estimates (β)	t	p
C1 manner	Front-to-back	**[pt]~[ft]**	**[pt]: -0.293 [ft]: -0.717**	**4.042**	**0.0001**
	Back-to-front	**[kt]~[xt]**	**[kt]: 0.797 [xt]: 0.243**	**5.276**	**0.0001**
	Interaction of *Place order* and *C1 manner*:				
		plosive C1 and fricative C1		0.888	0.3747
C2 manner	Front-to-back	**[pt]~[ps]**	**[pt]: -0.260 [ps]: -0.508**	**2.212**	**0.0272**
		[pt]~[pl]	[pt]: -0.260 [pl]: -0.471	1.468	0.1424
		[ps]~[pl]	[ps]: -0.508 [pl]: -0.471	0.276	0.7826
	Back-to-front	**[kt]~[ks]**	**[kt]: 0.764 [ks]: 0.392**	**4.194**	**0.0001**
		[kt]~[kl]	[kt]: 0.764 [kl]: 0.644	1.246	0.2129
		[ks]~[kl]	**[ks]: 0.392 [kl]: 0.644**	**2.077**	**0.0381**
	Interaction of *Place order* and *C2 manner*:				
		plosive C2 and fricative C2		1.398	0.1625
		plosive C2 and lateral C2		0.162	0.8716
		fricative C2 and lateral C2		1.264	0.2065

Table 3.3: Paired comparisons and interactions testing Predictions 2b and 2p from the LMMs fit to z-scores of the intergestural lag durations, pooled across speakers, in sequences contrasting in *C1 manner* and those contrasting in *C2 manner*. Rows in **bold** indicate a level of significance of p<0.05.

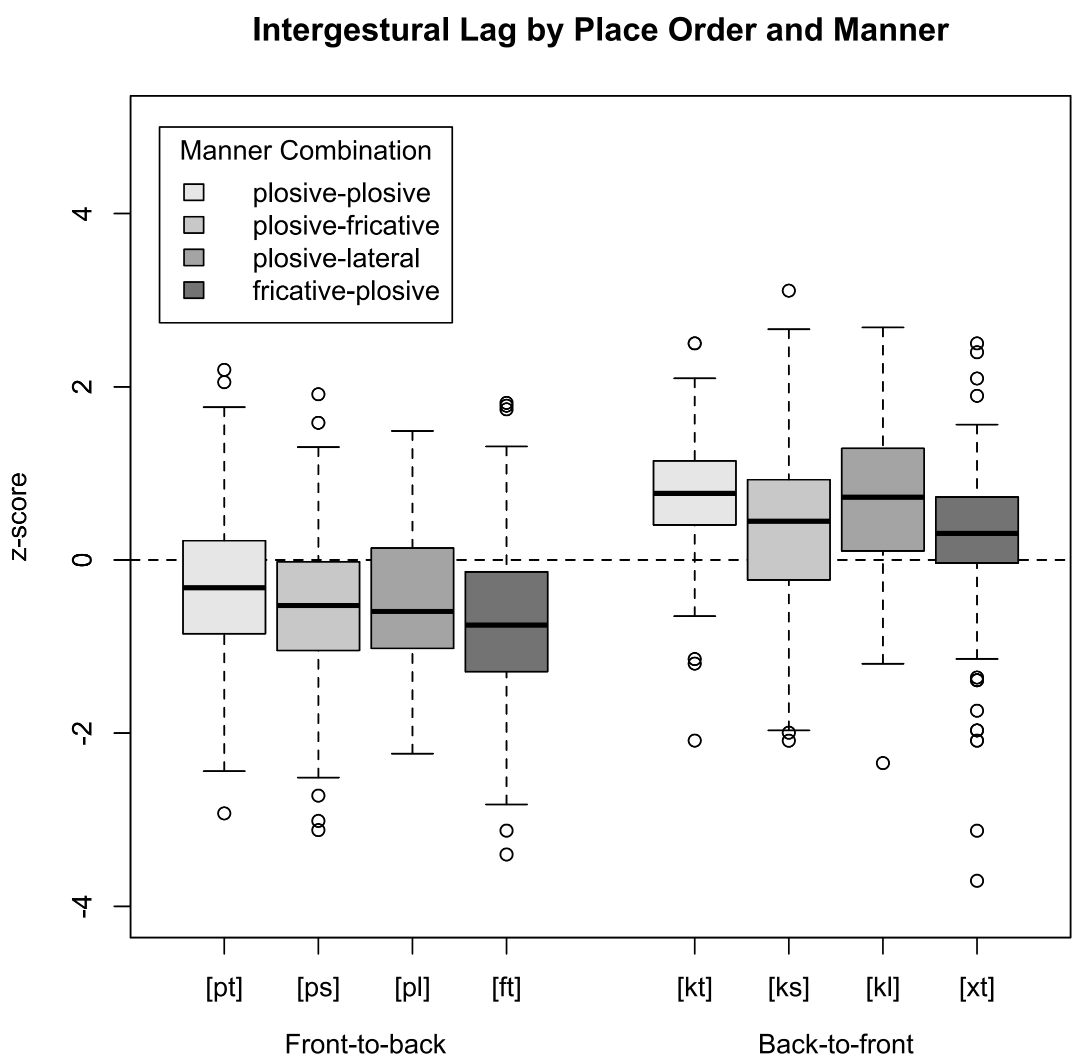


Figure 3.6: *z*-scores of intergestural lag durations for the CC sequences [pt ps pl ft kt ks kl xt], by place order and manner combination, across all speakers. Negative values do not necessarily indicate negative intergestural lag times, but rather below-average lag durations.

For *place order* and *C2 manner*, both predictions (2b and 2p) state that constriction manner should not systematically influence intergestural lag in front-to-back, labial-coronal [pt ps pl]. The LMM computed for these effects showed that, for these sequences, intergestural lag for plosive-plosive [pt] was significantly longer than for plosive-fricative [ps] (estimate: -0.507), but not significantly longer than that for plosive-lateral [pl] (estimate: -0.473). Lag measures for

[ps] and [pl] also did not differ significantly. For back-to-front, dorsal-coronal [kt ks kl xt], Prediction 2p stated that intergestural lag should be longer for [kt] than for the other manner combinations. This prediction was upheld for [kt]~[ks] (estimates of 0.797 and 0.392, respectively) and [kt]~[xt] (estimate: 0.243), but not for [kt]~[kl] (estimate: 0.644). Additionally, the difference in lag for [ks] and for [kl]—longer lag in [kl] than in [ks]—was significant. Contrary to Prediction 2p, there was no significant interaction between *place order* and *C2 manner*.

Because a substantial amount of the total variance in the two models was attributable to the random effect of *speaker* (*place order* and *C1 manner*: 37.6%: *place order* and *C2 manner*: 37.7%), individual-speaker patterns were assessed by performing separate LMM analyses on the data for each speaker. The remainder of this chapter discusses the results of these LMMs, organized by findings for front-to-back, labial-coronal sequences (§3.4.1) and those for back-to-front, dorsal-coronal sequences (§3.4.2). Discussions of manner effects in these sections examine whether individual speakers produced positive or negative lag, since the validity of the predictions for perceptual recoverability (Prediction 2p) depends on whether speakers produce overlap between the constriction intervals of C1 and C2. Comparisons for C1 manner (plosive, fricative) and C2 manner (plosive, fricative, lateral) corresponding to the individual-speaker LMMs are presented in Figures 3.7–3.10, with the results for each speaker reported in the Appendix, §B.5–B.9.

3.4.1 Front-to-back, labial-coronal sequences

Findings for the effect of *C1 manner* on intergestural coordination for each speaker are shown in Figure 3.7, and the statistical results are reported in the Appendix, §B.5. Speakers generally appear to produce longer lag in plosive-plosive [pt] than in fricative-plosive [ft] in their intergestural lag estimates, but this effect was significant only for speakers S05, S07, and S10. The differences occurred regardless of whether speakers produced only positive lag (S07) or extensive negative lag (S05 and S10). However, two speakers' production behaviors appeared to deviate from the pattern, with speaker S06's and speaker S08's intergestural lag estimates for [ft] being longer than those for [pt], although these differences were not statistically significant. These two speakers produced only positive lag for these sequences.

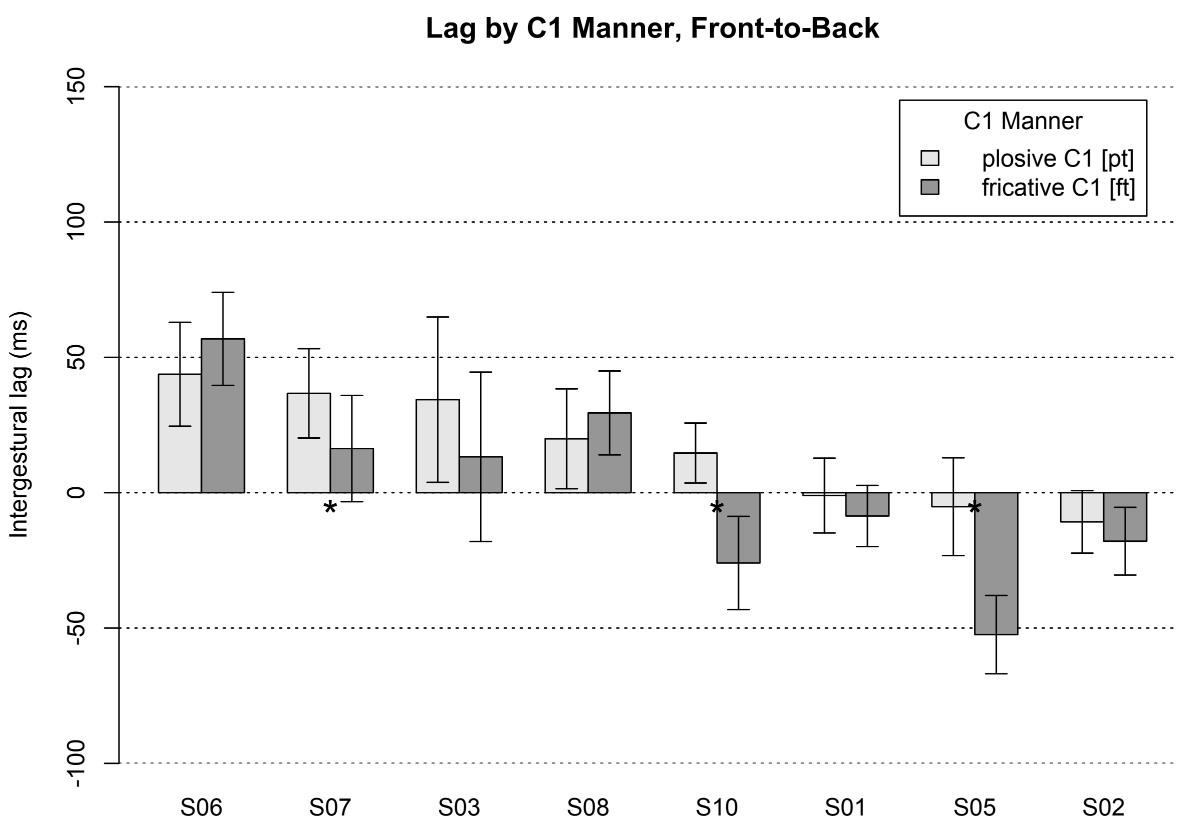


Figure 3.7: Intergestural lag durations (ms) for front-to-back sequences [pt ft], for each speaker. Speakers are ordered from longest to shortest mean lag duration in [pt] sequences. Asterisks signify a significant effect of place order, at $p<0.05$. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

The findings for the effect of *C2 manner* on front-to-back, labial-coronal sequences [pt ps pl] are shown in Figure 3.8 and in the Appendix, §B.7. These data also exhibit considerable variation among speakers. For two speakers (S05, S10), plosive-plosive [pt] had significantly longer lag than plosive-lateral [pl] (S05: -5.2 ms in [pt], -29.1 ms in [pl], $p = 0.0060$; S10: 14.7 ms in [pt], -14.5 ms in [pl], $p = 0.0001$), and, for speaker S07, this difference was marginal (36.71 ms in [pt], 17.61 ms in [pl], $p = 0.0500$). Speaker S10 also produced significantly longer lag in plosive-fricative [ps] than in plosive-lateral [pl] sequences (11.0 ms in [ps], -14.5 ms in [pl], $p = 0.0006$), whereas for one other speaker (S01) this pattern was reversed (-5.1 ms in [ps], 12.8 ms in [pl], $p = 0.0428$). Notably, three of these four speakers produced negative lag, although the lag values for some of the sequences were predominantly positive (S07: [pt], [ps], and [pl]; S10: [pt] and [ps]; S01: [pl]). The difference between lag durations in plosive-plosive [pt] and plosive-fricative [ps] was not significant for any speaker, even though lag estimates for [ps] were shorter than those for [pt] for all speakers. This weak [ps]~[pt] tendency held for speakers who rarely produced negative lag (S06, S07, S03, S08) as well as for those who persistently produced negative lag (S02, S05, S01, S10).

Speakers S01, S03, and S06 showed a tendency to produce relatively long intergestural lag in plosive-lateral clusters compared to other [p]-initial clusters and compared to most other speakers. Although their productions of [pl] did not involve significantly longer lag than those of [pt] or [ps], this superficial pattern of later lateral-C2 achievement relative to C1 for these speakers also emerges in the data for their back-to-front, dorsal-coronal sequences [kt ks kl] (discussed in the next section, §3.4.2). In terms of the range of their lag values, speakers S03 and S06 produced both predominantly positive lag, while speaker S01 produced substantial amounts of negative lag.

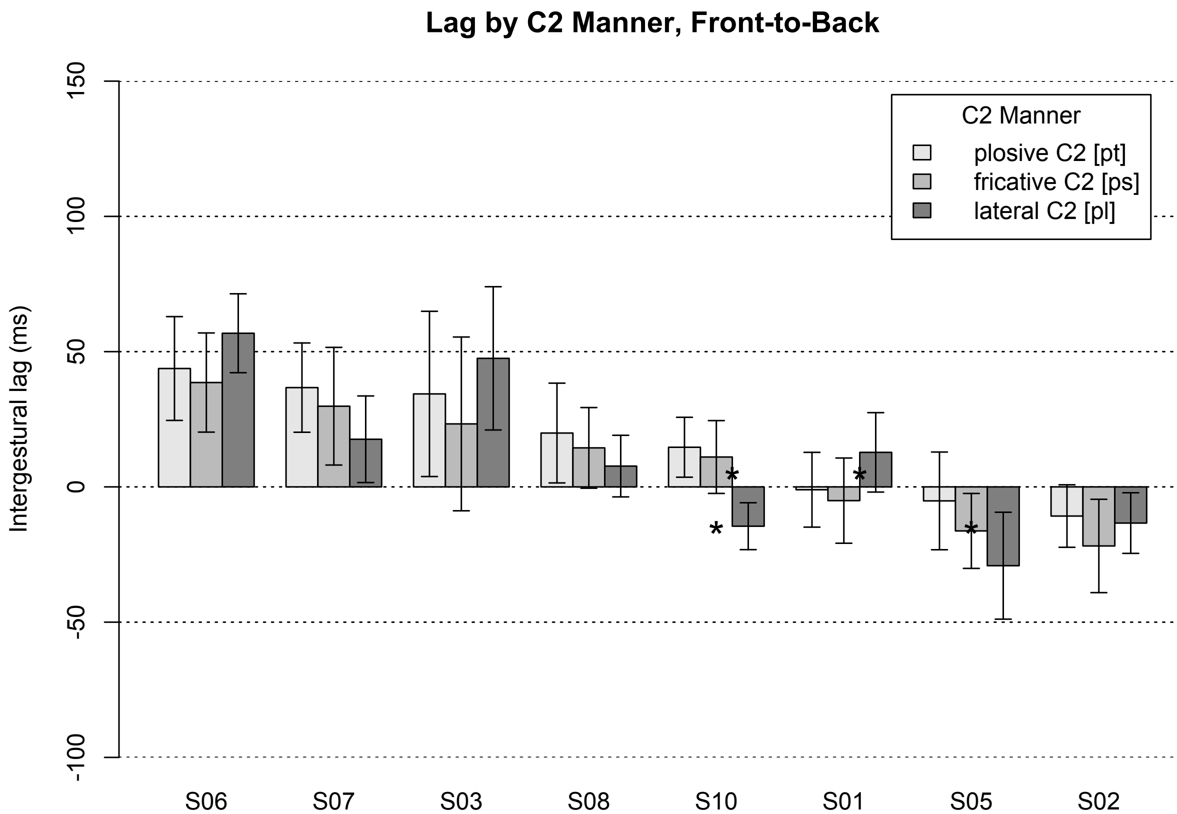


Figure 3.8: Intergestural lag durations (ms) by C2 manner for [pt ps pl], for each speaker. Asterisks positioned above zero indicate significant differences (p<0.05) between [pt]~[ps] (between bars for [pt] and [ps]) and [ps]~[pl] (between bars for [ps] and [pl]), while asterisks positioned below zero a significant difference between [pt]~[pl]. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

Summary: In the analyses for individual speakers, intergestural lag was significantly longer in plosive-plosive [pt] than in fricative-plosive [ft] sequences for three of the eight speakers, with the remaining five speakers' lag estimates conforming to the same pattern. Lag in plosive-plosive [pt] sequences was not significantly longer than that in plosive-fricative [ps] for any individual speaker. However, because each speaker's lag estimates followed the expected pattern of greater lag in [pt] clusters than in [ps] clusters, the C2 manner effect was significant in the general LMM for *place order* and *C2 manner* across speakers.

3.4.2 Back-to-front, dorsal-coronal sequences

Overall, intergestural lag values for the back-to-front, dorsal-coronal sequences [kt ks kl xt] tended to be positive and large across speakers, although one speaker (S01) frequently produced negative lag (constriction overlap) in [xt] sequences. Results for individual-speaker LMMs with the fixed effect *C1 manner* are provided in Figure 3.9 and in the Appendix, §B.6. The effect of C1 manner in back-to-front (dorsal-coronal) sequences was significant for six speakers (S01, S02, S06–S10). Although not significant for the remaining two speakers (S03, S05), differences in lag estimates for [kt] and [xt] followed the direction of the effect produced by the others.

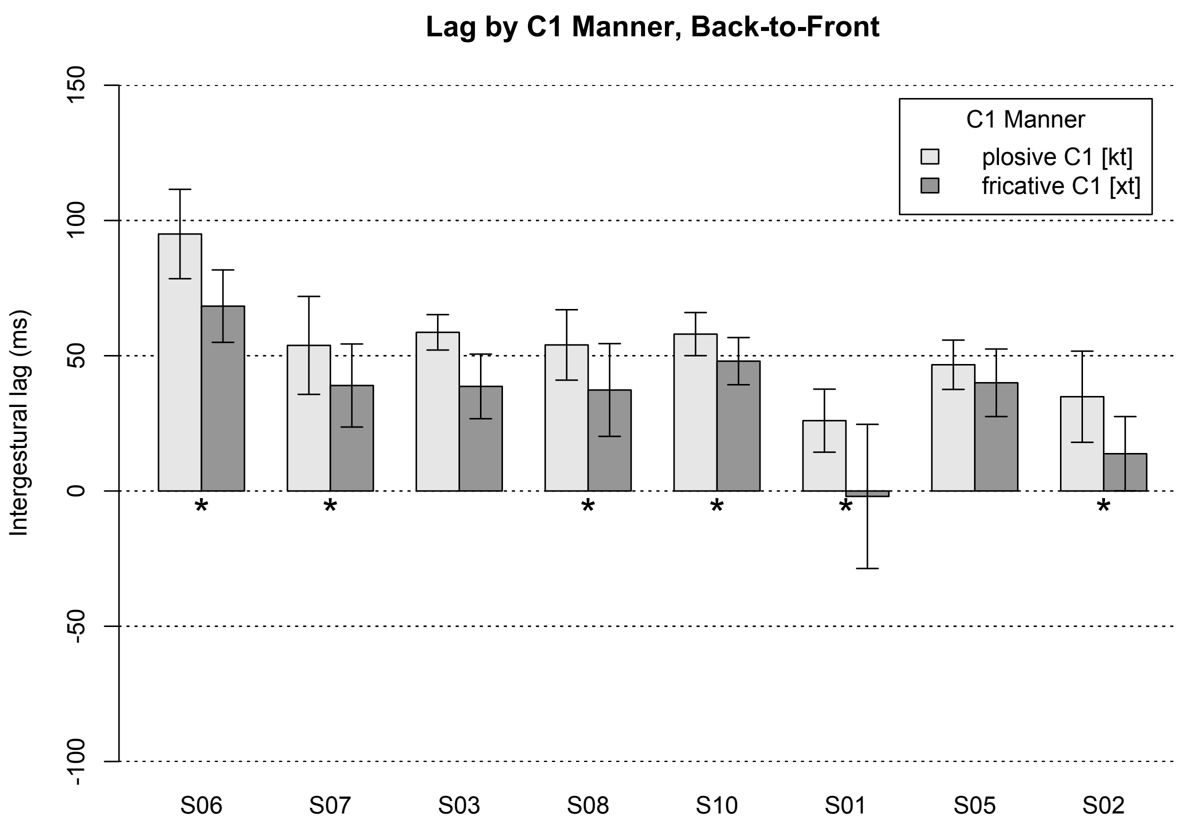


Figure 3.9: Intergestural lag durations (ms) for back-to-front sequences [kt xt], for each speaker. For consistency with Figures 3.7 and 3.8, speakers are ordered from longest to shortest lag duration in [pt] sequences. Asterisks signify a significant effect of place order, at $p<0.05$. Negative lag values indicate that the time of C2 achievement preceded the time of C1 release.

Figure 3.10 and the Appendix, §B.8, give the results for the effect of *C2 manner* in back-to-front (dorsal-coronal) [kt ks kl] clusters. As with the front-to-back (labial-coronal) sequences, some of the effects that were significant in the LMM conducted across speakers were not significant in results for individual speakers. Specifically, the across-speaker pattern of intergestural lag being longer in plosive-plosive [kt] than in plosive-fricative [ks] was significant only for speakers S02, S03, and S05 (S02: p = 0.0338; S03: p = 0.0002; S05: p = 0.0001). For three speakers, S01, S03, and S05, lag in plosive-lateral [kl] was significantly longer than that in plosive-fricative [ks] (S01: p = 0.0171; S03: p = 0.0015; S05: p = 0.0253). Finally, for speaker S08, lag in plosive-plosive [kt] was significantly longer than that in plosive-lateral [kl] (p = 0.0039).

These results are inconsistent with the biomechanical prediction (Prediction 2b), under which it was hypothesized that dorsal-coronal sequences would not show overall manner differences. However, as explained in Section §2.4, this prediction was tentative, as it assumed that the lingual demands of [kt], [ks], and [kl] constrictions were all comparable in these Greek clusters. However, if, for example, the plosive [t] constriction in [kt] takes longer to complete than the fricative [s] constriction in [ks] and/or the lateral [l] constriction [kl], then this pattern of shorter lag in [ks] than in [kt] and [kl] could be consistent with a biomechanical hypothesis. To address this possibility, these speakers' tongue contours—in addition to the lag patterns reported here—for the relevant sequences need to be examined (see further discussion in Section §4.2.2).

These findings for the effect of *C2 manner* in back-to-front (dorsal-coronal) sequences are also not fully consistent with the predictions made by a perceptual recoverability approach to intergestural timing (Prediction 2p). That approach was interpreted as predicting that lag in [kt] should be longer than that in both [ks] and [kl]. However, only the effect of longer lag in [kt] than in [ks] was observed, and this effect is significant for only three speakers (S02, S03, S05).

The predicted effect of longer lag in [kt] than in [kl] was only observed for one speaker (S08). While perceptual recoverability would seem to predict that lag duration in [ks] and [kl] should be similar, three speakers produced significantly longer lag in [kl] than in [ks] (S01, S03, S05).

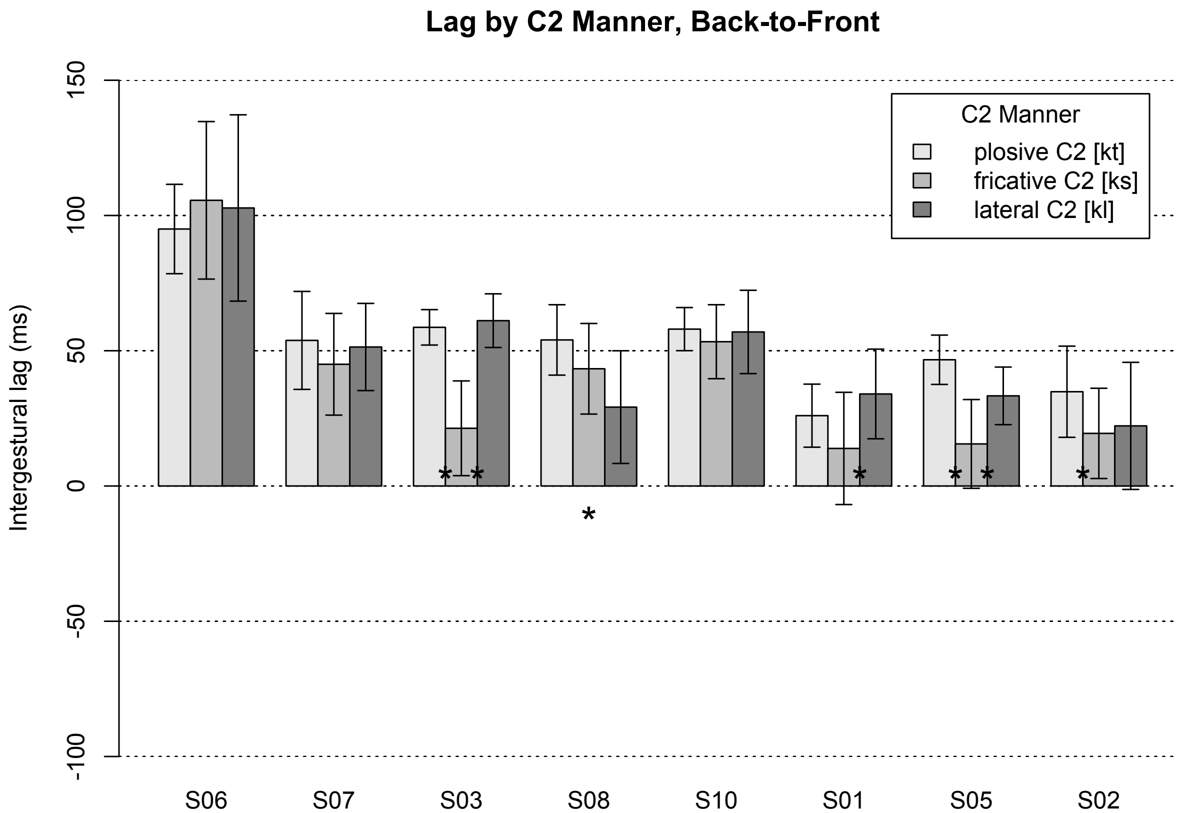


Figure 3.10: Intergestural lag durations (ms) by C2 manner for [pt ps pl], for each speaker. Asterisks situated above zero indicate significant differences (p<0.05) between [kt]~[ks] (between bars for [kt] and [ks]) and [ks]~[kl] (between bars for [ks] and [kl]), while asterisks situated below zero a significant difference between [kt]~[kl].

Of the three speakers who had longer intergestural lag estimates in plosive-lateral [pl] clusters than in plosive-plosive [pt] and plosive-fricative [ps] clusters (§3.4.1), two of them also exhibited significantly longer lag in plosive-lateral [kl] clusters than in plosive-fricative [ks] clusters: speaker S01 (p = 0.0171) and speaker S03 (p = 0.0015). The third speaker, S06, had a longer average lag duration for [kl] (102.8 ms) than for [kt] (95.0 ms), but not for [ks] (105.6 ms), and no difference in lag among the [kt], [ks], and [kl] contexts was significant for this speaker. Speaker S05 also produced significantly longer lag in [kl] (33.3 ms) relative to that in [ks] (15.6 ms), although this pattern does not parallel his productions of sequences with laterals in the front-to-back context (i.e., longer lag durations for [ps] than for [pl]). These observations indicate that Greek speakers vary, perhaps idiosyncratically, in terms of how they produce lateral gestures in plosive-lateral [pl kl] sequences. (This point is discussed further in Section §4.2.2.)

Summary: As with the front-to-back, labial-coronal comparisons, plosive-plosive [kt] sequences were produced with longer intergestural lag than both fricative-plosive [xt] sequences and plosive-fricative [ks] sequences. In addition, plosive-lateral [kl] sequences were generally produced with longer lag than plosive-fricative [ks] sequences. Many speakers deviated from the pattern of longer lag in [kt kl] than in [ks]: speaker S08 produced longer lag in [kt], relative to that in [kl], and three other speakers (S06, S07, S10), produced no significant differences in lag in back-to-front, dorsal-coronal sequences when C2 was a plosive, a fricative, or a lateral.

3.4.3 Interactions between place order and C1/C2 manner

Results for the interactions between the effects of *place order* and *C1 manner* and between the effects are *place order* and *C2 manner* for each speaker are reported the Appendix, §B.9. For five of the eight speakers, *place order* interacted with *C1 manner* (S01: p = 0.0475, S05: p = 0.0003, S06: p = 0.0001, S08: p = 0.0078, S10: p = 0.0001). For two of these speakers (S05, S10), lag was consistently longer in plosive-C1 than in fricative-C1 sequences, but the effect of *C1 manner* was greater in front-to-back, labial-coronal (Figure 3.7) than back-to-front, dorsal-coronal (Figure 3.9) sequences. In the productions of the three remaining speakers (S01, S06, S08), for front-to-back, labial-coronal [pt ft], lag durations in plosive-C1 and fricative-C1 sequences were not significantly different, and for back-to-front, dorsal-coronal [kt xt], lag was significantly longer when C1 was a plosive than when it was a fricative.

Results for the interaction between *place order* and *C2 manner* show that for two speakers (S05, S10), the effect of *C2 manner* on intergestural lag (Figures 3.8 and 3.10) was influenced by *place order* (S05: $p = 0.0270$ in [pt kt ps ks], $p = 0.0067$ in [ps ks pl kl]; S10: $p = 0.0018$ in [pt kt pl kl], $p = 0.0013$ in [ps ks pl kl]). For speaker S05, when C2 was a fricative, lag was especially short in back-to-front, dorsal-coronal sequences ([ks] < [kt kl]) but not in front-to-back, labial-coronal sequences ([ps] ≮ [pt pl]). For speaker S10, a different pattern arose: when the place order was front-to-back (labial-coronal), lag was shorter when C2 was a lateral than a plosive or fricative ([pl] < [pt ps]), while in back-to-front, dorsal-coronal sequences, C2 manner did not influence lag ([kl] ≮ [kt ks]).

3.5 Summary of Results

A summary of the significant comparisons for each speaker is presented in Table 3.4 for effects of order of place of articulation, Table 3.5 for effects of C1 manner and C2 manner, and Table 3.6 for the interactions between place order and C1 or C2 manner. The LMMs pooled across speakers, which used *z*-score values of intergestural lag, identified patterns in the group data that did not reach the designated 0.05 level of significance within individual speakers' LMM analyses. For example, the pooled C1 manner effect ([pt kt] > [ft xt]) was significant for roughly half of the speakers, and the pooled C2 manner effect ([pt kt] > [ps ks]) was rarely significant for individual speakers. However, the pooled and individual differences are unsurprising: for individual speakers, each test condition contained at most 25 tokens of intergestural lag (5 words × 5 repetitions), whereas the pooled analyses were performed across a much larger set of observations per condition.

	Place Order			
	[kt] > [pt]	[ks] > [ps]	[kl] > [pl]	[xt] > [ft]
S01	X	X	X	
S02	X	X	X	X
S03	X			
S05	X	X	X	X
S06	X	X	X	
S07	X		X	X
S08	X	X		
S10	X	X	X	X

Table 3.4: Summary of results for individual speakers, where "X" indicates a significant effect of place order on lag duration. The "greater than" symbol (>) indicates that lag durations for the CC sequence to the left were significantly longer than those for the CC sequence to the right.

Place order: The pooled results showed that speakers produced greater intergestural lag in plosive-plosive [kt] than [pt], plosive-fricative [ks] than [ps], plosive-lateral [kl] than [pl], and fricative-plosive [xt] than [ft]. Although these patterns were not significant for all speakers, the productions of many speakers who did not show a significant difference between the two place orders for these sequences exhibited a trend in the expected direction. These results initially appear to be more consistent with the biomechanical prediction for the effect of order of place of articulation (Prediction 1b) because perceptual masking (Prediction 1p) would seem to make a relatively weak contribution to place order patterns other than [kt]~[pt]. However, Section §4.2.1 offers a tentative explanation as to how perceptual-recoverability factors might have brought about the place-order pattern across speakers' CC sequence productions.

	C1 Manner		C2 Manner					
	[pt] > [ft]	[kt] > [xt]	[pt] > [ps]	[pt] > [pl]	[pl] > [ps]	[kt] > [ks]	[kt] > [kl]	[kl] > [ks]
S01		X			X			X
S02		X				X		
S03						X		X
S05	X			X		X		X
S06		X						
S07	X	X		(marginal)				
S08		X					X	
S10	X	X		X	[pl] < [ps]			

Table 3.5: Summary of results for individual speakers, where "X" indicates a significant effect of C1 or C2 manner on lag duration. The "greater than" symbol (>) indicates that lag durations for the CC sequence to the left were significantly longer than those for the CC sequence to the right, while the "less than" symbol (<) indicates an effect in the opposite direction.

	Place Order * C1 Manner	Place Order * C2 Manner
S01	X	
S02		
S03		
S05	X	X
S06	X	
S07		
S08	X	
S10	X	X

Table 3.6: Summary of results for individual speakers, where "X" indicates a significant interaction between place order and either C1 or C2 manner.

C1 manner: The general LMM found a main effect of C1 manner in both place-order contexts, with greater lag in plosive-plosive sequences ([pt], [kt]) than in corresponding fricative-plosive sequences ([ft], [xt]). An effect of C1 manner on the labial-coronal [pt]~[ft] sequences is consistent with avoidance of perceptual masking (Prediction 2p), but not with the biomechanical interpretation assumed here (Prediction 2b).

C2 manner: For the effect of C2 manner, the general LMM revealed, for both place orders, significantly longer intergestural lag in plosive-plosive sequences ([pt] ,[kt]) than in plosive-fricative ([ps], [ks]) sequences, and significantly longer lag in plosive-lateral sequences than in plosive-fricative sequences for the back-to-front, dorsal-coronal context only (i.e., [kl] > [ks]). Similar to C1-manner effects, the finding of longer lag when C2 is a plosive than when it is a fricative for labial-coronal sequences [pt]~[ps] is not predicted by a biomechanical account (Prediction 2b). (See Section §4.2.2 for a more detailed discussion.)

In comparisons between plosive-lateral and plosive-fricative contexts, intergestural lag was longer in [kl] than in [ks], whereas lag was not longer in [pl] than in [ps]. This pattern may support the biomechanical prediction (1b), provided that there is evidence that the demands on the motion of the tongue for the constrictions for [t] and for [s] differ when the gestures are coordinated with a preceding dorsal [k] C1 constriction. This potential outcome is explored in greater detail in Section §4.2.2.

The perceptual-recoverability and biomechanical interpretations of this study's results are discussed in Chapter IV.

CHAPTER IV

Discussion and Conclusion

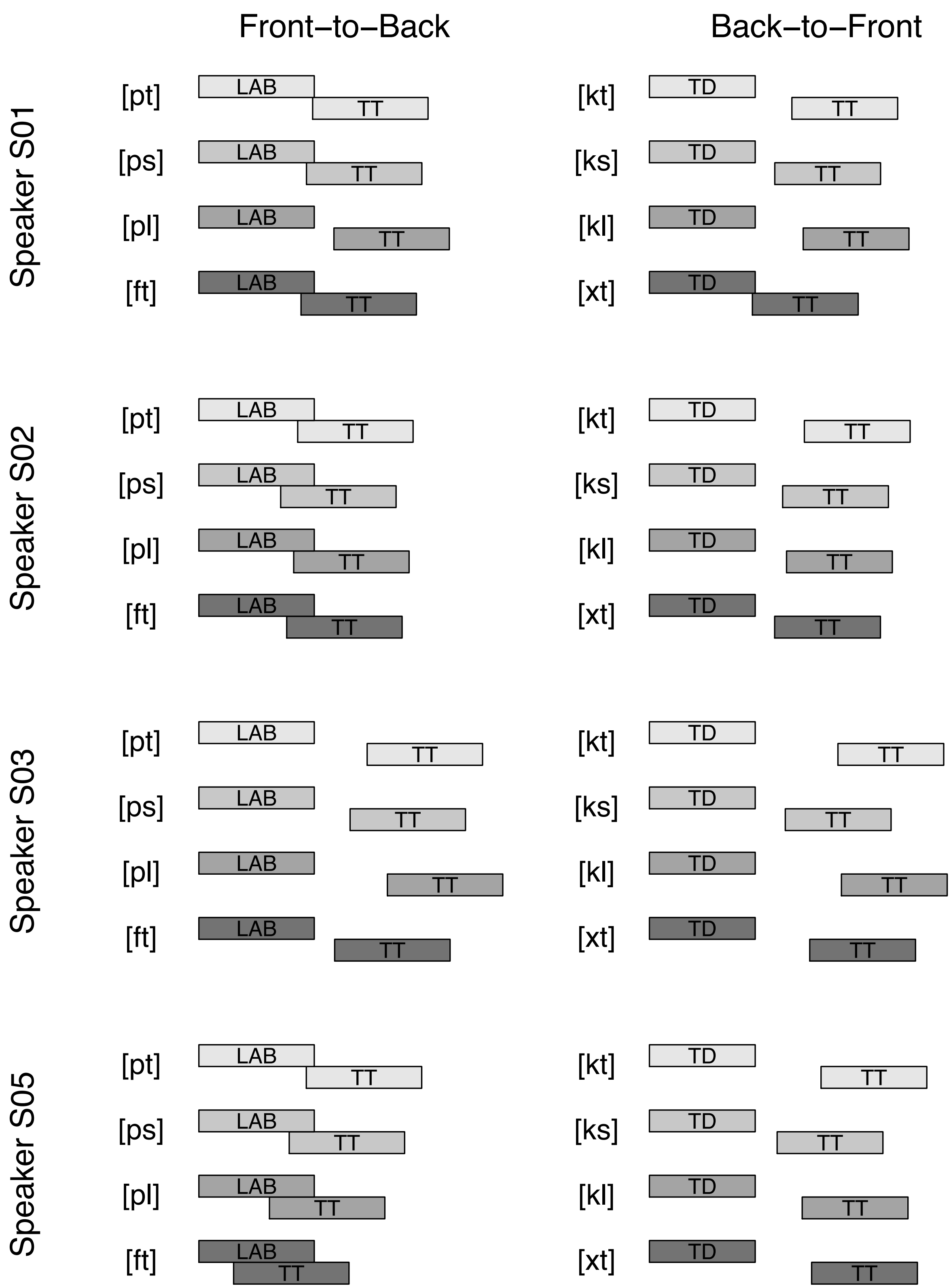
In this experiment, native speakers of Modern Greek read sentences containing words that began with the clusters [pt ps pl ft kt ks kl xt]. Speakers' tongues and lips were imaged with ultrasound and camera, and labial, tongue-tip, and tongue-dorsum apertures were tracked and measured over time. This study examined the influences of place order, C1 manner, and C2 manner on gestural coordination during these CC sequences, and an investigation of how these phonetic effects interact yielded a richer picture than has been presented in the literature as to how biomechanical and perceptual-recoverability pressures might guide production.

Place Order		C1 Manner	C2 Manner
[pt] < [kt]	[pl] < [kl]	[ft] < [pt]	[ps] < [pt], but [ps] ≮ [pl]
[ps] < [ks]	[ft] < [xt]	[xt] < [kt]	[ks] < [kt] and [ks] < [kl]

Table 4.1: Summary of pooled *z*-score results across speakers. The "less than" symbol (<) indicates that the *z*-scores for the CC sequence to the left were significantly smaller than those for the CC sequence to the right.

The results from Chapter III, summarized in Table 4.1 across all speakers, reveal that the place-order effect—i.e., shorter lag in front-to-back (labial-coronal) than in back-to-front (dorsal-coronal) sequences—was robust across speakers. For place order and C1 manner, the pooled mixed linear model found the same effect of C1 manner in both place orders (i.e., no interaction between these effects), with shorter lag when C1 was a fricative than when it was a plosive. The model also did not reveal a significant overall interaction between place order and C2 manner, although comparisons between lag in C2-manner pairs showed longer intergestural lag in [kl] than in [ks] and no difference in the duration of intergestural lag in [pl] and [ps]. For both place orders, intergestural lag was longer when C2 was a plosive ([pt], [kt]) than when it was the corresponding fricative ([ps], [ks]).

Individual-speaker results (shown in Tables 3.4–3.6) were generally consistent with effects revealed by the pooled analysis. However, these effects were not always significant—or even trends—at the individual level. Figure 4.1 presents schematic diagrams for the intergestural lag estimates from the mixed linear models for individual speakers. Individual patterns in these results are discussed in more detail in the following sections, in the context of assessing the theoretical approaches to gestural coordination.


Front–to–Back
Back–to–Front
Speaker S01
Speaker S02
Speaker S03
Speaker S05
[pt]
[ps]
[pl]
[ft]
[kt]
[ks]
[kl]
[xt]
LAB
TT
TD

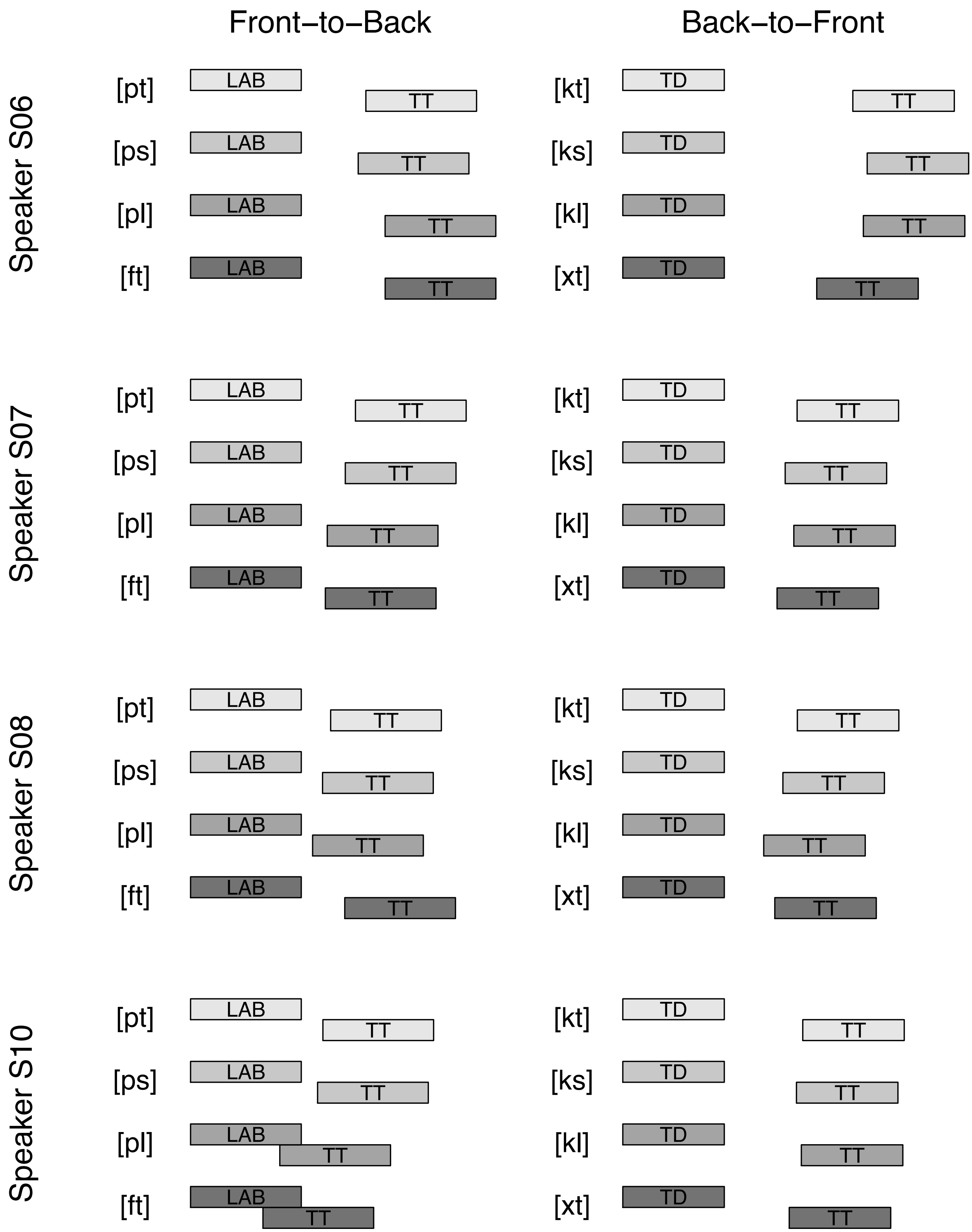


Figure 4.1: Schematization of results for speakers S01–S10, based on lag duration estimates in the mixed linear models for individual speakers.

4.1 Review of Predictions

In assessing the level of support that this study's results provide for the two hypotheses about articulator coordination introduced in Chapter I, I reiterate the hypothetical predictions presented in Section §2.4:

Speech Biomechanics: Speakers' productions are necessarily governed to some extent by the physiological capabilities of the vocal tract. These capabilities can be expected to constrain consecutive articulatory gestures, especially those that involve interdependent articulators such as the tongue dorsum and tongue tip. In particular, the lingual gestures required for the (back-to-front) dorsal-coronal sequences [kt ks kl xt] might be expected to have relatively long temporal separation. If instead the articulators required for the gestural constrictions are independent, as in (front-to-back) labial-coronal sequences [pt ps pl ft], then speakers might be expected to produce more overlap for the constriction gestures. Broadly speaking, the articulator(s) involved in forming vocal-tract constrictions might be expected to more heavily influence gestural timing than does constriction degree (but see Section §4.2.2). Consequently, this theoretical perspective does not yield general predictions concerning changes in timing of gestural coordination due to changes to C1 or C2 manner.

Perceptual Recoverability: If speakers' productions are planned so as to meet the perceptual needs of the listener, then effects of both constriction location (place) and degree (manner) are expected to emerge. Due to possible acoustic masking, back-to-front (dorsal-coronal) sequences should exhibit less overlap than front-to-back (labial-coronal) sequences. However, because acoustic masking is especially likely in overlapping back-to-front plosive-plosive sequences (here, [kt]), C2 manner should influence gestural timing to a greater degree in back-to-front (dorsal-coronal) than in front-to-back (labial-coronal) contexts. For similar perceptual reasons, speakers should produce longer lag when the first consonant is a plosive, which depends on its release cues for accurate perception, than when it is a fricative, which provides critical acoustic information during constriction.

These predictions are schematized in Figure 4.2. In the sections that follow, I discuss the extent to which the results of this study support or refute these sets of predictions for gestural coordination in word-initial CC clusters.

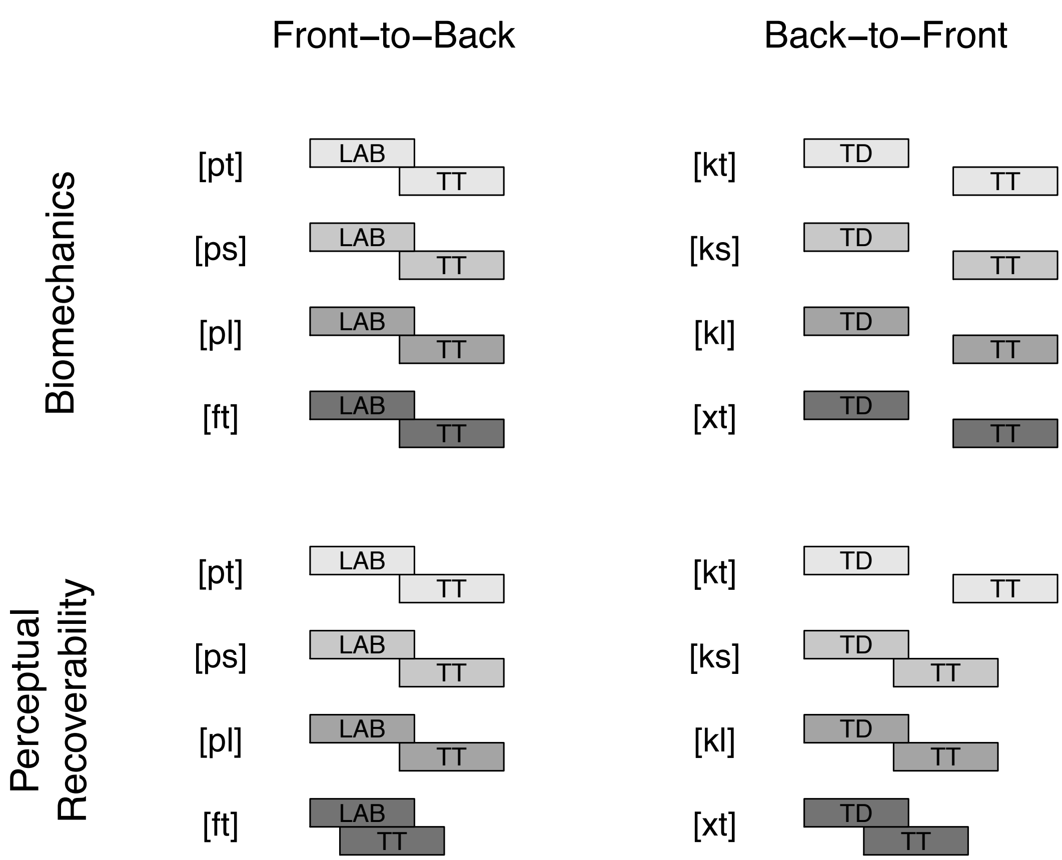


Figure 4.2: Schematization of the predictions made in Section §2.4 and summarized in Section §4.1. To facilitate comparisons with Figure 4.1, this figure presents gestures as constriction intervals over time, whereas the corresponding Figure 2.17 presents the same gestures as degrees of aperture for the relevant articulators.

4.2 Overall Strength of the Theoretical Approaches

4.2.1 Perceptual Recoverability

The findings for intergestural lag in [pt] and [kt] sequences appear to support the perceptual-recoverability prediction that speakers should produce longer lag for the onset CC sequences in which complete acoustic masking during gestural overlap is more likely. Lag durations in back-to-front (dorsal-coronal) [kt] contexts were about 40 ms longer than those in front-to-back (labial-coronal) [pt] contexts. However, recall from the discussion of predictions for place order and C1/C2 manner (Section §3.4) that the set of predictions for perceptual recoverability apply especially to sequences in which the constriction intervals of C1 and C2 overlap—in this study, when the intergestural lag measures are negative. The fact that productions of all speakers exhibited, on average, longer lag in [kt] than in [pt], regardless of whether their productions contained any negative intergestural lag (i.e., overlap), possibly weakens the interpretation that recovery of essential acoustic information motivates the place-order effect. Because the productions of speakers whose C1 releases rarely overlapped with C2 achievement (S03, S06, S07) still showed greater separation of [kt] than [pt] gestures, it seems doubtful that these speakers produced longer lag in [kt] in order to maintain the audibility of the release of C1 [k]. That is, for these speakers, the achievement of the C2 constriction was never sufficiently early to possibly mask the release of the C1.

However, it is possible that the place-order pattern was originally motivated by perceptibility in the plosive-plosive sequences alone, and this pattern was extended to non-plosive–plosive and plosive–non-plosive sequences, yielding systematic timing differences between front-to-back and back-to-front even when acoustic recoverability was not in jeopardy. Consistent with this perspective, for the speakers in this study, the place-order effect was most reliably present in the production of plosive-plosive sequences [pt kt], and less reliably present for the other manner combinations. That is, if perceptibility primarily drives the place-order effect for plosive-plosive sequences, then it is not unexpected for speakers to produce the place-order pattern less consistently for sequences in which C1 or C2 is not a plosive (i.e., is a fricative or lateral).

Gestural timing has often been approached with an assumption that gestural overlap is pervasive, and that masking due to articulatory coordination is a serious possibility. These findings for Greek initial CC sequences show that this is not always the case: during the production of CC clusters, some (Modern Greek) speakers occasionally produced gestural overlap between the two consonants' constriction intervals, while other speakers did not coordinate their articulators in ways that created this kind of intergestural overlap. Although the production data do not seem to strongly support a perceptual-recoverability motivation for the place-order patterns across this range of speakers' productions, acoustic/perceptual data for these Greek clusters are needed to address the question of whether the differences in gestural timing might increase perceptibility under conditions of both negative and positive intergestural lag as measured here. Investigating this production-perception relationship could take the form of an acoustic analysis of the production data collected in this study, or a perceptual experiment that tests the effect of these gestural-overlap patterns on phonetic recoverability by listeners.

Stronger evidence of the possible influence of constraints on gestural coordination imposed by perceptual recoverability comes from the intergestural-lag comparisons across C1 manners. The finding of longer lag in plosive-plosive [pt kt] than in fricative-plosive [ft xt], with a greater degree of overlap in fricative-plosive sequences, conforms to the perceptually-oriented prediction for the effect of C1 manner. While only two speaker's (S07, S10) productions exhibited the C1-manner effect in both front-to-back and back-to-front contexts at a statistically significant level, the analysis pooled across speakers showed this pattern, for both place orders and across all speakers, at a level of statistical significance. The main effect of C1 manner may be due to differences in the availability of acoustic cues to plosives versus fricatives, as discussed in Section §2.4.1. If the constriction intervals of C1 and C2 coincide, and C1 is a plosive, some degree of acoustic masking is possible, even in a front-to-back context, since pressure build-up behind the C1 closure could be insufficient for an audible release if the oral airstream is blocked by an early C2 constriction behind it. If the avoidance of the masking of release cues is indeed a factor underlying lengthened lag intervals in plosive-plosive [pt kt] sequences relative to those in fricative-plosive [ft xt] sequences across speakers, this lag effect is consistent with an interpretation according to which speakers use their perceptual knowledge in the planning of articulatory timing to maintain perceptual recoverability.

While the C1-manner effect on lag can be explained in perceptual terms, the effect of C2 manner identified in Chapter III (and visually represented in Figure 4.1) seems less clearly motivated by recoverability. The expected influence of C2 manner—less overlap for plosive-plosive than for other CC sequences—held across speakers for both place orders when lag in plosive-plosive was compared with that in plosive-fricative (lag durations for [ps] < [pt] and [ks] < [kt]); two speakers also exhibited the expected pattern when lag in plosive-plosive [pt] was compared with that in plosive-lateral [pl] ([pl] < [pt]). Although the two speakers who produced significantly less overlap in [pt] than [pl] did indeed produce overlapping labial and coronal constrictions for [pl], the three speakers who produced significant [ks] < [kt] lag duration patterns did not have overlapping dorsal and coronal constrictions for these sequences (as assessed by my overlap measure). Presumably, then, there should have been little if any acoustic masking of C1 by C2, yet these speakers nonetheless exhibited an effect of C2 manner. (Again, these expectations merit verification against the acoustic signal and/or perceptual testing.)

Speaker S10 produced lag patterns that would appear to more clearly support a perceptual-recoverability explanation: C2 manner influenced lag durations in front-to-back contexts but not in back-to-front contexts. Moreover, this speaker produced negative lag, i.e., gestural overlap, in 52.3% of front-to-back CC sequences compared to 0% of back-to-front and exhibited no manner differences among back-to-front sequences [kt ks kl xt] and increasing intervals of overlap among front-to-back sequences [pt ps pl ft], as the likelihood of acoustic masking during overlap decreased. Speaker S10's production behaviors are consistent with perceptual recoverability, because, for this speaker, any perceptually-oriented timing adjustments should occur primarily in front-to-back contexts, in which gestural overlap was actually present.

4.2.2 Speech Biomechanics

The amount of intergestural lag produced by individual speakers varied dramatically, from lag durations as short as -100 ms for speaker S01 to as long +191 ms for speaker S06. The effect of which articulators were involved—lips and tongue tip or tongue tip and tongue dorsum—was robust across speakers, regardless of how much lag each speaker produced. The biomechanical explanation introduced in Section §1.4.1 proposes that articulatory independence/interdependence is a main reason behind the difference in articulatory timing between place-order contexts. Figure 4.3, which presents frame-by-frame contour data extracted from single-token productions of front-to-back, labial-coronal [pt] (top row) versus back-to-front, dorsal-coronal [kt] (bottom row) by speaker S03, demonstrates how the achievement of tongue-tip constriction for a coronal C2 [t] can occur synchronously with with the formation labial C1 [p] closure, but coincides with release of the dorsal C1 [k] closure. In [kt], the tongue body first moves to the posterior region of the oral cavity to form the velar constriction, but moves forward to complete the coronal C2 constriction, a forward motion that continues even after tongue tip contact is achieved. The lag between [k] and [t] closures can be viewed as the result of the time that it takes for the tongue body to complete this anterior movement after the velar constriction has been released.

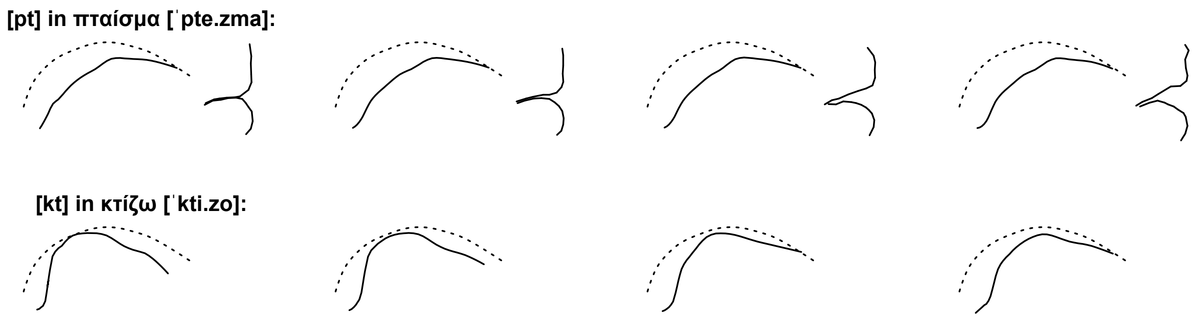


Figure 4.3: Contour traces of the surface of the tongue ([pt] and [kt]) and of the upper and lower lips ([pt] only) extracted from a single token for each of the plosive-plosive sequences [pt] and [kt] over four consecutive ultrasound and camera frames (16.7 ms apart = 60 fps), as produced by speaker S03. When C1 was labial, constriction at the tongue tip for C2 was already complete at the moment C1 release, but when C1 was dorsal, achievement of same tongue tip constriction was delayed by approximately 33 ms.

These kinematic observations underscore the importance of biomechanical relationships between the vocal-tract articulators during production, as suggested by findings from Mooshammer *et al.* (1995) and Kühnert *et al.* (2006) (discussed in Section §1.4). In the current study, the same "looping" motion of the tongue described by Mooshammer *et al.* can be seen across speakers' (back-to-front) dorsal-coronal [kt] productions, as shown in Figure 4.4. During the transition from [k] to [t], speakers initially raise the tongue body to form a velar constriction for [k] and subsequently move the tongue body forward and/or downward to form a coronal constriction for [t]. Since the tongue is restricted by this kind of motion, the constriction for [t] will tend to follow the release of the constriction for [k]. This constraint on the motion of the tongue body out of velar constrictions and into coronal constrictions may explain why intergestural lag in (back-to-front) dorsal-coronal [kt] was longer—and almost always positive (see Figures 3.9, 3.10, and 4.1)—than in (front-to-back) labial-coronal [pt].

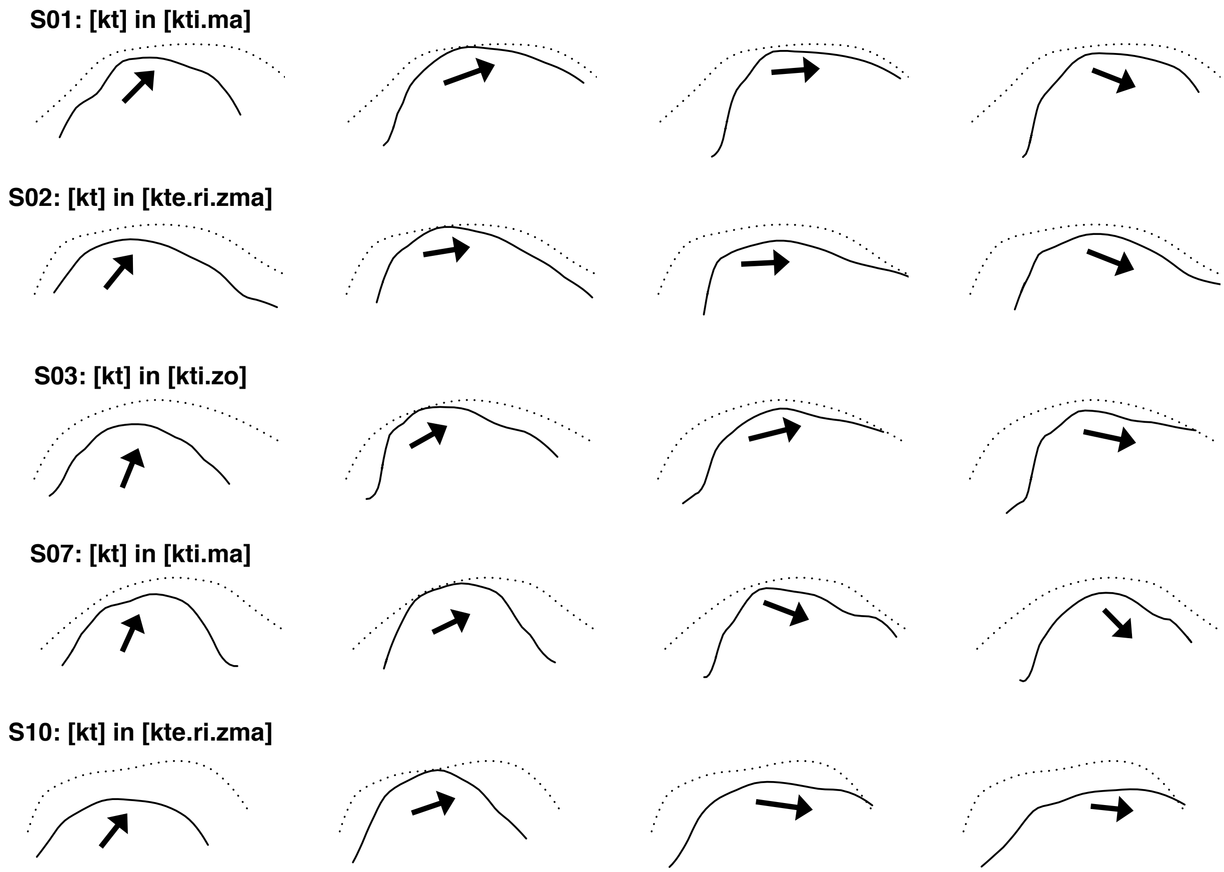


Figure 4.4: Contours extracted from individual tokens of [kt] for five speakers from this study (S01, S02, S03, S07, S10), depicting the transition from a dorsal C1 to a coronal C2. The frame rates vary between 10 and 15 fps, depending on the speaker, in order to capture as much of the motion of the tongue as possible. Arrows indicate the approximate direction of tongue-body movement in each frame.

However, the looping characterization as a persistent cause of longer lag in dorsal-coronal than in corresponding labial- coronal sequences cannot be entirely accurate, because, as is evident in the data from some speakers' plosive–non-plosive and non-plosive–plosive sequence productions, lag in dorsal-coronal was not always significantly longer than lag in labial-coronal: [ks] > [ps] only for six speakers; [kl] > [pl] only for six speakers; [xt] > [ft] only for four speakers. Nonetheless, almost without exception, speakers did not produce overlapped dorsal and coronal constrictions (as assessed by this study's C1 release – C2 achievement measure; see Figures 3.9 and 3.10).

The biomechanical approach, as interpreted here, does not predict a systematic effect of C1 manner on gestural timing in [xt] compared to [kt] and [ft] compared to [pt] sequences. Even if the dorsal-to-coronal lingual constraints differed for [xt] and [kt], there is no expectation for coordination patterns to differ for the labial-coronal sequences. Thus, the finding that constriction gestures were less overlapped for fricative-plosive than for plosive-plosive sequences regardless of the articulators involved is unexpected in the biomechanical approach. A perceptually-oriented approach appears to provide a better explanation for influences of C1 manner on C2 timing (for reasons presented in Section §4.2.1).

For reasons similar to those just described, any differences in intergestural lag across C2 manners should also be negligible. Nonetheless, C2 manner was shown to influence gestural timing irrespective of the articulators involved: 1) longer lag when C2 is a plosive than when it is a fricative ([pt] > [ps]; [kt] > ks]), and 2) longer lag when C2 is a lateral than when it is a fricative ([kl] > [ks]). Here again, the labial-coronal pattern would appear to be more consistent with a perceptual account. However, as discussed in Sections §1.6 and §2.4.2, differences in the coordination of dorsal-coronal sequences contrasting in C2 manner might be grounded in articulatory factors governing coordination of multiple lingual gestures.

In terms of the gestural-overlap differences between [kt] and [ks] sequences, of interest is whether, and how, the demands on tongue movement differ for these sequences. As an initial step in addressing this issue, the productions of three speakers who varied in the duration of intergestural lag in [ks] relative to that in [kt] (S03, S06, S07) were inspected in detail. As shown in Figures 4.5–4.7, none of these speakers' productions exhibit clear, systematic differences in the location of dorsal constriction for [k] in [kt] versus [ks]. Moreover, speaker S03's productions (Figure 4.5) exhibit only relatively small differences in the shape of the tongue configuration as the [t] constriction is formed compared to formation of the [s] constriction. By contrast, in speaker S06's productions (Figure 4.6), the midsagittal tongue contour in [kt] differs markedly from that in [ks]: as the dorsal constriction begins to release, the surface of the tongue assumes a concave ("saddle") shape into a plosive [t] constriction but not into a fricative [s] constriction. Speaker S07's productions (Figure 4.7), on the other hand, show a difference both in the shape of the tongue contour and the location along the palate at which the coronal constriction is made, with a much wider constriction for [s], which begins behind the alveolar ridge. Moreover, throughout the transition from [k] to [s], speaker S07's tongue maintains a convex ("domed") shape, as if the gestures for these two sounds were co-produced. Although these observations do not explain the timing patterns for [kt] compared to [ks], they provide an indication of the types of additional analyses relevant to further addressing the role of constraints on tongue motion in lingual-lingual sequences.

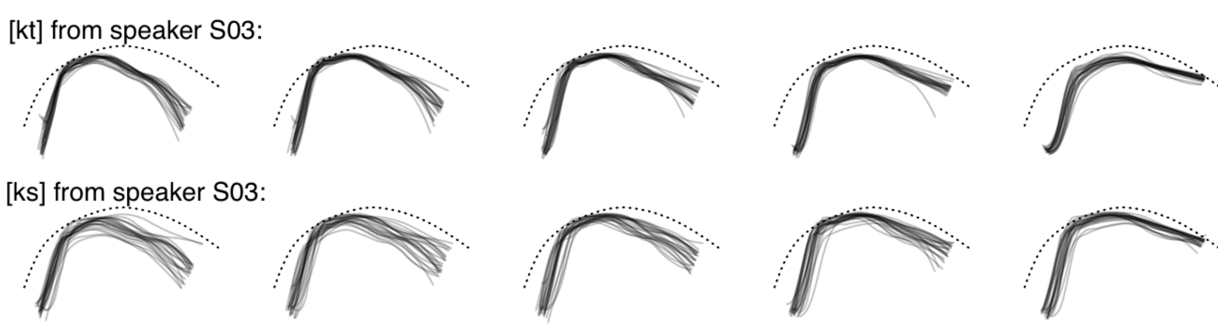


Figure 4.5: Aggregated tongue contours from all productions of [kt] (top row) and [ks] (bottom row) by speaker S03 (25 tokens each). The tongue contours move out of a dorsal [k] constriction in the first frames of each sequence and into a coronal [t] or [s] constriction by the last frame of the sequences. (Frames are 16.7 ms apart = 60 fps.)

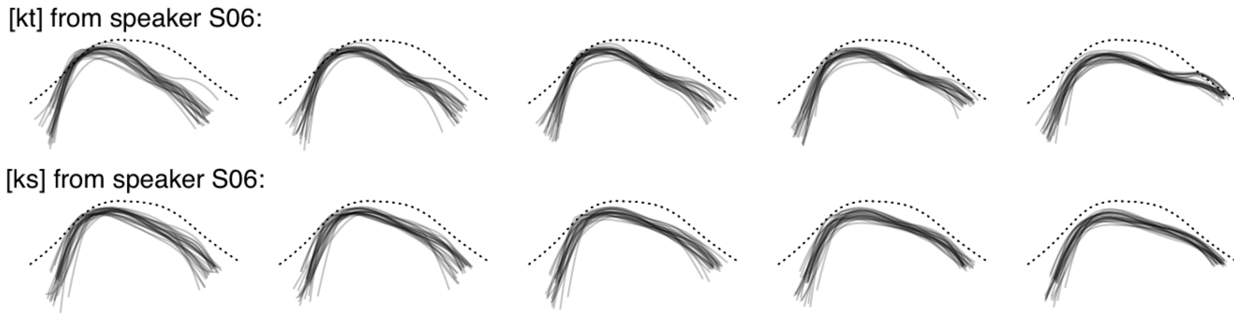


Figure 4.6: Aggregated tongue contours from all productions of [kt] (top row) and [ks] (bottom row) by speaker S06 (25 tokens each). The tongue contours move out of a dorsal [k] constriction in the first frames of each sequence and into a coronal [t] or [s] constriction by the last frame of the sequences. (Frames are 16.7 ms apart = 60 fps.)

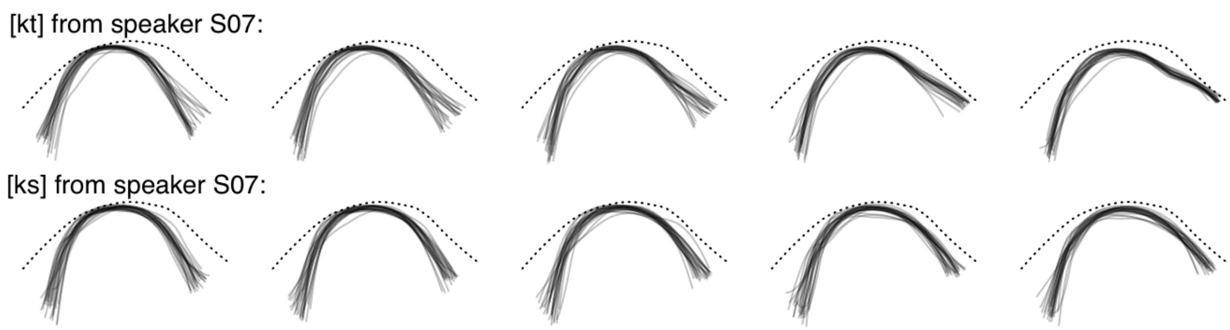


Figure 4.7: Aggregated tongue contours from all productions of [kt] (top row) and [ks] (bottom row) by speaker S07 (25 tokens each). The tongue contours move out of a dorsal [k] constriction in the first frames of each sequence and into a coronal [t] or [s] constriction by the last frame of the sequences. (Frames are 16.7 ms apart = 60 fps.)

Regarding the earlier timing of C2 constriction relative to C1 constriction in plosive-fricative [ps ks] sequences, it may be that this pattern is neither biomechanically nor perceptually motivated, but may rather have its source in the special status of these CC sequences in Modern Greek as being written by the single Greek letters *ψ* [psi] and *ξ* [ksi], both in word-initial and word-medial contexts.[2] If at some level of abstract representation, the sounds in these clusters are more closely associated with each other than those in all other clusters in this study, which are each orthographically represented with a sequence of two letters, there may be a relatively tight coordination between the two gestures in [ps] and [ks] at the stage of phonetic planning. If speakers coordinate onset [ps] and [ks] as if they are singletons, this may result in greater simultaneity between the gestural movements for C1 and for C2 and, correspondingly, in shorter lag between the C1 and C2 constriction intervals.

As reported in Sections §3.4.1 (results for front-to-back, labial-coronal sequences) and §3.4.2 (results for back-to-front, dorsal-coronal sequences), individual speakers showed different timing patterns for the coronal lateral constriction for [l] in plosive-lateral sequences, particularly in labial-coronal (front-to-back) [pl]. Previous research has suggested that the variability in the articulation of lateral approximants may be due, at least in part, to the fact that laterals often involve the coordination of two lingual gestures: a primary tongue-tip constriction in the coronal region and a secondary dorsum-raising gesture toward the soft palate (Sproat & Fujimura, 1993; Gick *et al.*, 2006). Thus, variation in how lateral approximants are articulated stems from differences in the temporal alignment of the two component gestures across speakers and prosodic contexts (Proctor, 2009; Lin, 2011).

The productions of plosive [t] and lateral [l] by speaker S03 (Figure 4.8), who tended to produce late lateral C2 achievement relative to C1 (based on lag values computed by the LMM), can be compared with corresponding productions by speakers S07 (Figure 4.9) and S10 (Figure

[2] In only a small handful of Greek words (less than ten items found in *GreekLex*, Ktori *et al.*, 2008), word-internal [ks] is written as two separate letters (*κσ*), but only in the context of long consonantal sequences across morphological boundaries (e.g., [e**k-s**ka.ˈfi] *ε**κσ**καφή* ‘excavation’, [ˈe**k-s**ta.si] *έ**κσ**ταση* ‘ecstacy’, [e**k-s**tɾa.ˈti.a] *εκ**στ**ρατεία* ‘campaign, expedition’) and in the single loanword [ˈgaŋ**k.s**teɾ] *γκάγ**κσ**τερ* ‘hoodlum, gangster’. No items for which word-internal [ps] was spelled as two separate letters (*πσ*) were found.

4.10), who tended to produce earlier lateral C2 achievement in their intergestural lag measures (also based on LMM lag values). Contour traces extracted from tokens of [pt] and [pl] reveal that the shape of the tongue during speaker S03's lateral productions are distinctly different from those of the other speakers. Before the completion of tongue-tip raising to form a coronal constriction for [l], speaker S03 raises his tongue dorsum toward the back of the oral cavity (forming the "saddle" contour characteristic of velarized laterals). In order to complete the coronal constriction for [l], this speaker must move his tongue body forward so that the raised tongue-tip can reach the alveodental region. This anterior movement presumably contributes to the roughly 50-ms delay in the achievement of full coronal constriction after release of the labial constriction for [p]. Speaker S03's lateral production behavior may explain why intergestural lag in his productions of plosive-laterals [pl kl] tended to be longer than lag in other manner sequences, although a more rigorous comparison of each of the tongue shapes in his lateral and non-lateral productions is needed to confirm this account.

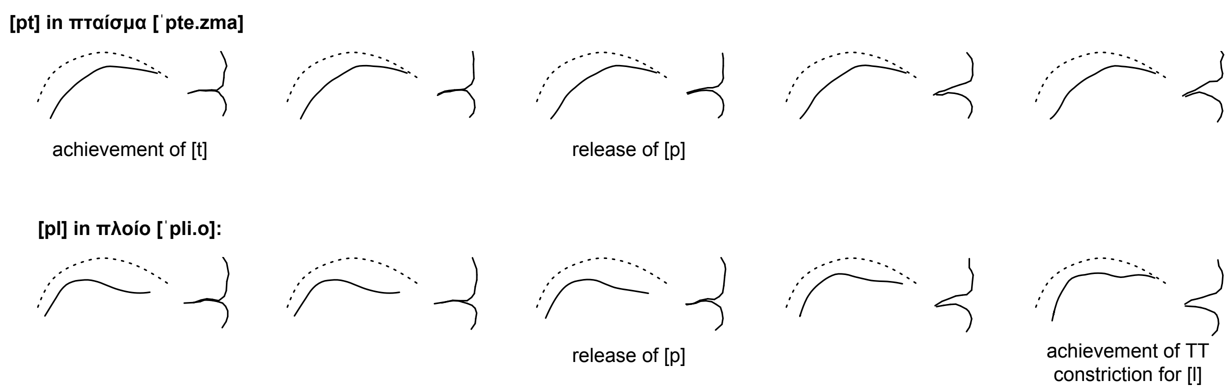


Figure 4.8: Contour traces of the tongue and lips extracted from single tokens of [pt] and [pl], as produced by speaker S03. Successive frames were 16.7 ms apart (60 fps).

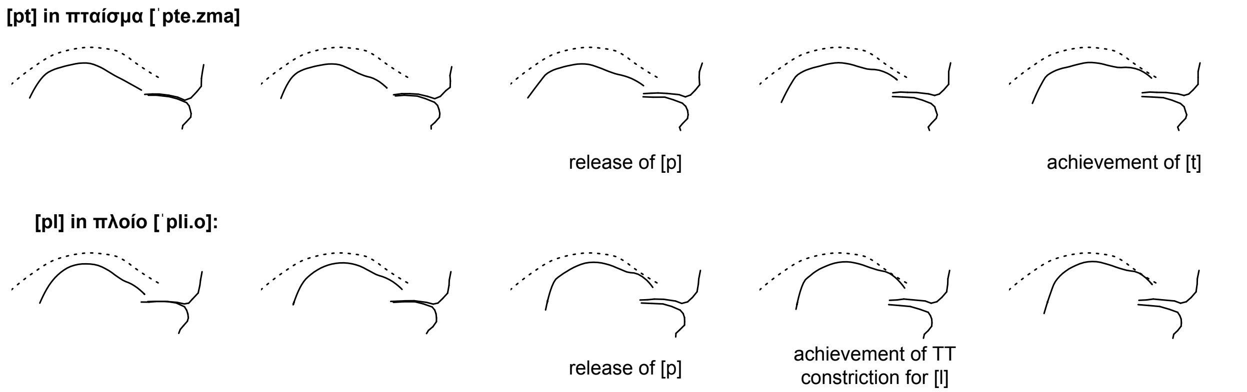


Figure 4.9: Contour traces of the tongue and lips extracted from single tokens of [pt] and [pl], as produced by speaker S07. Successive frames were 16.7 ms apart (60 fps).

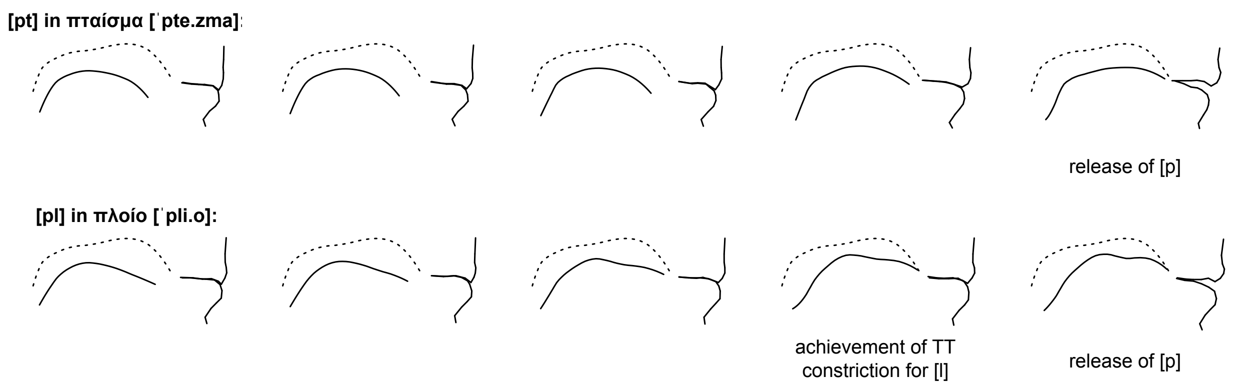


Figure 4.10: Contour traces of the tongue and lips extracted from single tokens of [pt] and [pl], as produced by speaker S10. Successive frames were 16.7 ms apart (60 fps).

While some degree of articulatory variation across different speakers' laterals is expected, this study's results show differences according to speaker in how laterals are coordinated with preceding plosive constriction gestures. Interestingly, the pattern of longer lag durations in plosive-lateral [pl kl] sequences only occurred in the production of the oldest male speakers in the study. More extensive testing on a larger group of both male and female Greek speakers would be needed to verify whether the velarization of Modern Greek laterals is a matter of sociolinguistic variation.

4.3 Conclusion

The data collected in this study are suggestive of both biomechanical and perceptual-recoverability factors' playing a role in determining how articulatory gestures are coordinated in consonant-consonant sequences. The effect of place order, already well-established in the literature, was also robust in these data and was shown to be consistent with a biomechanical basis, according to which the primary cause of longer lag in lingual-lingual ("back-to-front") sequences [kt ks kl xt] than in labial-lingual ("front-to-back") sequences [pt ps pl ft] is the greater degree of articulatory interdependence in the former. In contrast, the effect of C1 manner on gestural timing provided support for the role of perceptual recoverability: arguably, in order to avoid the masking of release cues to plosive C1, the timing of the C1 constriction in plosive-plosive [pt kt] was delayed relative to fricative-plosive [ft xt]. The place-independent effect of C2 manner, in which intergestural lag was longer in plosive-plosive [pt kt] than in plosive-fricative [ps ks], was not entirely consistent with either the biomechanical account or the perceptual-recoverability account, and this effect may instead be a consequence of the special status of these latter sequences as single Greek letters (*ψ* and *ξ*), which may have resulted in a temporally-close coordination between the two consonant gestures. Data from this study also show that speakers varied in the timing of C2 in plosive-lateral sequences, potentially due to differences between speakers in the shape and motion of the tongue during the production of the Greek lateral approximant [l]. While the articulatory behaviors of most speakers in this study showed no clear gestural timing patterns consistent with the avoidance of acoustic masking in back-to-front (dorsal-coronal) sequences, the productions of one speaker (S10) presented patterns of intergestural lag that suggest that the timing of that speaker's front-to-back (labial-coronal) plosive-plosive and plosive-fricative sequences may have been subject to perceptual enhancement. Future testing will be needed to verify whether such an articulatory strategy results in both a larger availability of acoustic cues in the auditory signal and improved perceptual recovery by the listener. Further analysis of the production data that takes advantage of the detailed information about lingual contours present in ultrasound video can provide additional insights into the relationship between the articulatory behaviors of the tongue and their acoustic consequences.

APPENDIX A

Experimental Stimuli

This section contains a complete list of the 80 target and 20 filler items used in the experiment, along with their IPA transcriptions, English glosses, syllable counts, and lexical frequencies. Although words were selected from the *GreekLex* corpus (Ktori *et al.*, 2008), reported frequencies are taken from the *Institute for Language & Speech Processing (ILSP) Psycho-Linguistic Resource* (IPLR) (Protopapas *et al.*, 2012). IPLR counts are preferred because *GreekLex* counts use only base-morpheme frequencies in a lemma database, while the IPLR counts take into consideration the lexical frequencies of each morphologically-derived form.

CC type	CC	V	IPA transcription	Greek orthography	English gloss	No. of syllables	*IPLR* frequency
Plosive+plosive	pt	i	[ˈ**pt**i.si]	πτήση	‘flight’	2	25.781
		e	[ˈ**pt**e.ɾi.ɣa]	πτέρυγα	‘wing’	3	14.447
		e	[ˈ**pt**e.zma]	πταίσμα	‘fault’	2	2.402
		e	[ˈ**pt**e.ɾaɾ.xos]	πτέραρχος	‘air chief marshal’	3	2.030
		i	[ˈ**pt**ixosi]	πτύχωση	‘plaiting, bunching’	3	0.237
	kt	i	[ˈ**kt**i.ma]	κτήμα	‘estate’	2	7.105
		i	[ˈ**kt**i.zma]	κτίσμα	‘building (n.)’	2	4.804
		i	[ˈ**kt**i.si]	κτήση	‘possession’	2	1.015
		i	[ˈ**kt**i.zo]	κτίζω	‘build (1.SG.PRES.)’	2	0.135
		e	[ˈ**kt**e.ɾi.zma]	κτέρισμα	‘grave offerings’	3	0.034

CC type	CC	V	IPA transcription	Greek orthography	English gloss	No. of syllables	*IPLR* frequency
Plosive + fricative	ps	i	[ˈ**psi**.fos]	ψήφος	‘vote (n.)’	2	13.262
		e	[ˈ**pse**.ma]	ψέμα	‘lie, untruth’	2	9.338
		i	[ˈ**psi**.xos]	ψύχος	‘cold (MASC.SG.)’	2	4.128
		e	[ˈ**pse**.ftis]	ψεύτης	‘liar, storyteller’	2	1.624
		i	[ˈ**psi**.no]	ψήνω	‘bake (1.SG.PRES.)’	2	0.101
	ks	e	[ˈ**kse**.ɾo]	ξέρω	‘know (1.SG.PRES.)’	2	132.997
		i	[ˈ**ksi**.lo]	ξύλο	‘wood’	2	20.401
		e	[ˈ**kse**.nos]	ξένος	‘foreign (MASC.SG.)’	2	16.510
		i	[ˈ**ksi**.fos]	ξίφος	‘sword, blade’	2	1.387
		e	[ˈ**kse**.ɾa]	ξέρα	‘reef, shoal’	2	0.406
Plosive + liquid	pl	e	[ˈ**ple**.on]	πλέον	‘over, more’	2	433.331
		i	[ˈ**pli**.o]	πλοίο	‘ship, vessel’	2	42.731
	pɾ	i	[ˈ**pɾi**.ka]	προίκα	‘dowry’	2	4.872
		e	[ˈ**pɾe**.za]	πρέζα	‘pinch (n.)’	2	1.049
		i	[ˈ**pɾi**.zo]	πρήζω	‘bloat (1.SG.PRES.)’	2	0.034
	kl	i	[ˈ**kli**.ma]	κλίμα	‘climate’	2	169.232
		e	[ˈ**kle**.vo]	κλέβω	‘steal (1.SG.PRES.)’	2	0.643
	kɾ	e	[ˈ**kɾe**.as]	κρέας	‘meat’	2	13.499
		i	[ˈ**kɾi**.a]	κρύα	‘cold (n.)’	2	4.940
		i	[ˈ**kɾi**.vo]	κρύβω	‘hide, cover (1.SG.PRES.)’	2	3.552
Fricative + plosive	ft	e	[ˈ**fte**.o]	φταίω	‘be at fault (1.SG.PRES.)’	2	2.571
		e	[ˈ**fte**ɾ.na]	φτέρνα	‘heel; counter’	2	0.677
		e	[ˈ**fte**.xtis]	φταίχτης	‘culprit’	2	0.338
		i	[ˈ**fti**.no]	φτύνω	‘cough up (1.SG.PRES.)’	2	0.203
		e	[ˈ**fte**.ɾi]	φτέρη	‘fern’	2	0.203
	xt	i	[ˈ**xti**.pi.ma]	χτύπημα	‘strike, bump (n.)’	3	25.273
		e	[ˈ**xte**.ni]	χτένι	‘scallop’	2	1.421
		e	[ˈ**xte**.ni.zma]	χτένισμα	‘hairdo, coif’	3	0.507
		i	[ˈ**xti**.pos]	χτύπος	‘clank, knock (n.)’	2	0.304
		i	[ˈ**xti**.zo]	χτίζω	‘construct (1.SG.PRES.)’	2	0.237

Table A.1: IPA transcriptions, Greek spellings, English glosses, syllable counts, and lexical frequencies of the 40 target words containing word-initial CC sequences in this study.

CC type	CC	V	IPA transcription	Greek orthography	English gloss	No. of syllables	*IPLR* frequency
Plosive + plosive	pt	e	[ɣɾa.ˈ**pt**es]	γραπτές	‘written (FEM.PL.)’	2	4.635
		e	[a.ˈ**pt**es]	απτές	‘tangible (FEM.PL.)’	2	0.812
		i	[a.na.ˈ**pt**i.so]	αναπτύσσω	‘develop (1.SG.PRES.)’	4	0.237
		i	[si.na.ˈ**pt**i]	συναπτή	‘consecutive (FEM.SG.)’	3	0.169
		i	[va.ˈ**pt**i.zo]	βαπτίζω	‘baptize (1.SG.PRES.)’	3	0.000
	kt	e	[tɾa.ˈ**kt**eɾ]	τρακτέρ	‘tractor’	2	15.022
		i	[a.ˈ**kt**i]	ακτή	‘coast’	2	11.638
		i	[ða.ˈ**kt**i.li.os]	δακτύλιος	‘ring (n.)’	4	2.470
		i	[va.ˈ**kt**i.ɾi.o]	βακτήριο	‘bacterium’	4	0.474
		e	[i.sa.ˈ**kt**eos]	εισακτέος	‘enrollment, entrant’	4	0.000
Plosive + fricative	ps	i	[çi.ɾa.ˈ**ps**i.a]	χειραψία	‘handshake’	4	3.992
		i	[a.ˈ**ps**i.ða]	αψίδα	‘arch, apse’	3	0.880
		i	[ta.ˈ**ps**i]	ταψί	‘pan’	2	0.643
		e	[ði.a.ˈ**ps**e.vðo]	διαψεύδω	‘refute (1.SG.PRES.)’	4	0.575
		i	[ka.ta.ˈ**ps**i.xo]	καταψύχω	‘deepfreeze (1.SG.PRES.)’	4	0.000
	ks	i	[ta.ˈ**ks**i]	ταξί	‘taxi’	2	23.886
		i	[ɣa.la.ˈ**ks**i.as]	γαλαξίας	‘galaxy’	4	2.301
		i	[e.fta.ˈ**ks**i.a]	ευταξία	‘orderliness’	4	1.015
		i	[a.ˈ**ks**i.zo]	αξίζω	‘deserve (1.SG.PRES.)’	3	0.913
		e	[a.ˈ**ks**e.xa.stos]	αξέχαστος	‘unforgettable (MASC.SG.)’	4	0.507
Plosive + liquid	pl	i	[a.ˈ**pl**i]	απλή	‘simple (FEM.SG.)’	2	70.609
		e	[ði.a.ˈ**pl**e.ko]	διαπλέκω	‘interweave (1.SG.PRES.)’	3	0.068
	pɾ	e	[a.ˈ**pɾ**e.pi.a]	απρέπεια	‘indecency’	4	1.319
		e	[ði.a.ˈ**pɾ**e.po]	διαπρέπω	‘excel (1.SG.PRES.)’	4	0.034
		i	[a.ˈ**pɾ**i.li.os]	Απρίλιος	‘April’	4	0.000
	kl	i	[ka.ta.ˈ**kl**i.ða]	κατακλείδα	‘conclusion, coda’	4	1.455
		i	[i.ɾa.ˈ**kl**is]	Ηρακλής	‘Hercules, Heracles’	3	0.000
	kɾ	i	[ma.ˈ**kɾ**is]	μακρύς	‘long (MASC.SG.)’	2	5.312
		e	[a.ˈ**kɾ**e.os]	ακραίος	‘extreme (MASC.SG.)’	3	1.319
		i	[a.ˈ**kɾ**i.ða]	ακρίδα	‘grasshopper’	3	0.169
Fricative + plosive	ft	e	[a.ˈ**ft**es]	αυτές	‘they (FEM.)’	2	737.826
		i	[a.ˈ**ft**i]	αυτί	‘ear’	2	9.778
		e	[ka.ˈ**ft**es]	καυτές	‘hot (FEM.PL.)’	2	1.319
		i	[na.ˈ**ft**i.a]	ναυτία	‘nausea’	3	1.150
		i	[ta.ˈ**ft**i.zo]	ταυτίζω	‘equate (1.SG.PRES.)’	3	0.135

CC type	CC	V	IPA transcription	Greek orthography	English gloss	No. of syllables	*IPLR* frequency
Fricative + plosive (cont.)	xt	i	[pe.ta.ˈ**xt**i]	πεταχτεί	‘jettisoned (PASS.PART.)’	3	2.233
		i	[a.ˈ**xt**i.ða]	αχτίδα	‘ray, beam’	3	1.218
		e	[aɾ.pa.ˈ**xt**es]	αρπαχτές	‘grabbed (FEM.PL.)’	3	0.609
		i	[tɾa.da.ˈ**xt**i]	τρανταχτοί	‘loud (MASC.PL.)’	3	0.000
		e	[fo.na.ˈ**xt**es]	φωναχτές	‘aloud (FEM.PL.)’	3	0.000

Table A.2: IPA transcriptions, Greek spellings, English glosses, syllable counts, and lexical frequencies of the 40 target words containing word-medial CC sequences in this study.

IPA transcription	Greek orthography	English gloss	No. of syllables	*IPLR* frequency
[ˈθe.si]	θέση	‘position, place’	2	773.148
[zo.ˈi]	ζωή	‘life’	2	397.840
[ˈma.tça]	μάτια	‘eyes’	2	141.827
[ˈti.pos]	τύπος	‘type; press’	2	72.808
[i.ˈʝi.a]	υγεία	‘health’	3	55.486
[ɾa.de.ˈvu]	ραντεβού	‘appointment; rendezvous’	3	41.682
[ɾi.ˈθmos]	ρυθμός	‘rhythm’	2	39.381
[ˈi.ʎos]	ήλιος	‘sun’	2	20.334
[ˈnu.me.ɾo]	νούμερο	‘number’	3	20.300
[o.ˈðos]	οδός	‘street, road’	2	18.574
[bu.ˈka.li]	μπουκάλι	‘bottle’	3	6.631
[o.ˈði.o]	ωδείο	‘(musical) conservatory’	3	5.413
[ˈza.xa.ɾi]	ζάχαρη	‘sugar’	3	5.143
[ˈsku.pa]	σκούπα	‘broom’	2	3.958
[ʝa.tɾi.ˈkos]	ιατρικός	‘medical; medicinal (MASC.SG.)’	3	2.504
[do.ˈma.ta]	ντομάτα	‘tomato’	3	2.301
[ˈli.ma]	λήμμα	‘lemma’	2	2.165
[ˈθa.la.ses]	θάλασσες	‘seas’	3	6.124
[ˈti.no]	τείνω	‘tend, incline (1.SG.PRES.)’	2	0.372
[ˈli.o]	λύω	‘unravel, undo (1.SG.PRES.)’	2	0.000

Table A.3: IPA transcriptions, Greek spellings, English glosses, syllable counts, and lexical frequencies of the 20 filler items used in this study.

APPENDIX B

Statistical Results for the Effects of Place Order, C1 Manner, and C2 Manner and Interactions for Individual Speakers

The tables and figures in this appendix summarize the results of the linear mixed-effects models (LMMs) for each speaker in the study. All statistics reported in this section pertain to individual speakers and are based on the three LMM-types for: 1) *place order* alone (§B.1–B.4), 2) *place order* and *C1 manner* (§B.5–B.6), and 3) *place order* and *C2 manner* (§B.7–B.8). All LMMs in this section include the sole random factor of *item*. Tests for interactions between *place order* and *C1 manner* and between *place order* and *C2 manne*r are additionally reported in §B.9. The measurements of intergestural lag reported in the tables of this appendix are as described in Section §2.3, with intergestural lag referring to the duration of the interval between the release of C1 and the achievement of C2 in the Modern Greek clusters [pt ps pl ft kt ks kl xt].

Speaker	Front-to-Back	Back-to-Front	*t*	*p*
S01	**[pt]: -1.05**	**[kt]: 26.00**	**5.01**	**0.0001**
S02	**[pt]: -10.78**	**[kt]: 34.85**	**7.33**	**0.0001**
S03	**[pt]: 34.37**	**[kt]: 58.65**	**2.60**	**0.0122**
S05	**[pt]: -5.17**	**[kt]: 46.67**	**6.97**	**0.0001**
S06	**[pt]: 43.76**	**[kt]: 95.00**	**6.78**	**0.0001**
S07	**[pt]: 36.70**	**[kt]: 53.82**	**2.32**	**0.0248**
S08	**[pt]: 19.91**	**[kt]: 54.00**	**5.06**	**0.0001**
S10	**[pt]: 14.66**	**[kt]: 58.00**	**10.64**	**0.0001**

Table B.1: Intergestural lag estimates (ms) of the two place orders for plosive+plosive sequences [pt kt], by speaker. **Bolded** data are significant at $p<0.05$.

Speaker	Front-to-Back	Back-to-Front	t	p
S01	**[ps]: -5.08**	**[ks]: 13.88**	**2.36**	**0.0228**
S02	**[ps]: -21.83**	**[ks]: 19.45**	**5.46**	**0.0001**
S03	[ps]: 23.28	[ks]: 21.33	0.18	0.8588
S05	**[ps]: -16.28**	**[ks]: 15.55**	**4.96**	**0.0001**
S06	**[ps]: 38.75**	**[ks]: 105.61**	**5.24**	**0.0001**
S07	[ps]: 29.83	[ks]: 45.00	1.77	0.0835
S08	**[ps]: 14.42**	**[ks]: 43.33**	**4.32**	**0.0001**
S10	**[ps]: 11.05**	**[ks]: 53.34**	**7.31**	**0.0001**

Table B.2: Intergestural lag estimates (ms) of the two place orders for plosive+fricative sequences [ps ks], by speaker. **Bolded** data are significant at $p<0.05$.

Speaker	Front-to-Back	Back-to-Front	t	p
S01	**[pl]: 12.77**	**[kl]: 34.02**	**2.23**	**0.0364**
S02	**[pl]: -13.37**	**[kl]: 22.22**	**3.06**	**0.0060**
S03	[pl]: 48.07	[kl]: 61.12	0.85	0.4038
S05	**[pl]: -29.13**	**[kl]: 33.32**	**6.46**	**0.0001**
S06	**[pl]: 56.80**	**[kl]: 102.78**	**2.85**	**0.0092**
S07	**[pl]: 17.61**	**[kl]: 51.39**	**3.45**	**0.0023**
S08	[pl]: 7.70	[kl]: 29.16	1.67	0.1105
S10	**[pl]: -14.53**	**[kl]: 56.95**	**9.39**	**0.0001**

Table B.3: Intergestural lag estimates (ms) of the two place orders for plosive+lateral sequences [pl kl], by speaker. **Bolded** data are significant at $p<0.05$.

Speaker	Front-to-Back	Back-to-Front	t	p
S01	[ft]: -8.61	[xt]: 6.61	0.76	0.4481
S02	**[ft]: -17.92**	**[xt]: 13.78**	**5.61**	**0.0001**
S03	[ft]: 13.26	[xt]: 38.67	1.58	0.1215
S05	**[ft]: -52.43**	**[xt]: 40.00**	**11.71**	**0.0001**
S06	[ft]: 56.83	[xt]: 68.33	1.77	0.0833
S07	**[ft]: 16.31**	**[xt]: 39.00**	**2.34**	**0.0235**
S08	[ft]: 29.47	[xt]: 37.33	1.14	0.2598
S10	**[ft]: -25.96**	**[xt]: 48.00**	**12.84**	**0.0001**

Table B.4: Intergestural lag estimates (ms) of the two place orders for fricative+plosive sequences [ft xt], by speaker. **Bolded** data are significant at $p<0.05$.

Speaker	Plosive C1	Fricative C1	*t*	*p*
S01	[pt]: -1.05	[ft]: -8.61	1.05	0.2966
S02	[pt]: -10.78	[ft]: -17.91	1.23	0.2202
S03	[pt]: 34.37	[ft]: 13.26	1.57	0.1191
S05	**[pt]: -5.17**	**[ft]: -52.43**	**6.16**	**0.0001**
S06	[pt]: 43.76	[ft]: 56.83	1.85	0.0668
S07	**[pt]: 36.71**	**[ft]: 16.31**	**2.77**	**0.0068**
S08	[pt]: 19.91	[ft]: 29.47	1.40	0.1642
S10	**[pt]: 14.66**	**[ft]: -25.96**	**8.14**	**0.0001**

Table B.5: C1 Manner—Intergestural lag estimates (ms) of C1 manner for front-to-back sequences [pt ft], by speaker. **Bolded** data are significant at p<0.05.

Speaker	Plosive C1	Fricative C1	*t*	*p*
S01	**[kt]: 26.00**	**[xt]: -2.00**	**3.89**	**0.0002**
S02	**[kt]: 34.85**	**[xt]: 13.78**	**3.46**	**0.0008**
S03	[kt]: 58.65	[xt]: 38.66	1.49	0.1398
S05	[kt]: 46.67	[xt]: 40.00	0.87	0.3872
S06	**[kt]: 95.00**	**[xt]: 68.33**	**3.78**	**0.0003**
S07	**[kt]: 53.82**	**[xt]: 39.00**	**1.99**	**0.0495**
S08	**[kt]: 54.00**	**[xt]: 37.33**	**2.44**	**0.0164**
S10	**[kt]: 58.00**	**[xt]: 48.00**	**2.00**	**0.0480**

Table B.6: C1 Manner—Intergestural lag estimates (ms) of C1 manner for back-to-front sequences [kt xt], by speaker. **Bolded** data are significant at p<0.05.

Speaker	Plosive C2	Fricative C2	Lateral C2	*t*	*p*
S01	[pt]: -1.05	[ps]: -5.07		0.59	0.5551
	[pt]: -1.05		[pl]: 12.77	1.67	0.0972
		[ps]: -5.07	**[pl]: 12.77**	**2.13**	**0.0353**
S02	[pt]: -10.78	[ps]: -21.83		1.54	0.1269
	[pt]: -10.78		[pl]: -13.37	0.30	0.7682
		[ps]: -21.83	[pl]: -13.37	0.94	0.3512
S03	[pt]: 34.37	[ps]: 23.28		1.13	0.2624
	[pt]: 34.37		[pl]: 47.52	1.04	0.2986
		[ps]: 23.28	[pl]: 47.52	1.92	0.0567
S05	[pt]: -5.17	[ps]: -16.28		1.76	0.0815
	[pt]: -5.17		**[pl]: -29.13**	**3.05**	**0.0028**
		[ps]: -16.28	[pl]: -29.13	1.64	0.1040
S06	[pt]: 43.76	[ps]: 38.58		0.54	0.5890
	[pt]: 43.76		[pl]: 56.80	1.11	0.2697
		[ps]: 38.58	[pl]: 56.80	1.54	0.1264
<u>S07</u>	[pt]: 36.71	[ps]: 29.83		0.89	0.3777
	<u>[pt]: 36.71</u>		<u>[pl]: 17.61</u>	<u>1.98</u>	<u>0.0500</u>
		[ps]: 29.83	[pl]: 17.61	1.27	0.2075
S08	[pt]: 19.91	[ps]: 14.42		0.81	0.4214
	[pt]: 19.91		[pl]: 7.69	1.40	0.1628
		[ps]: 14.42	[pl]: 7.69	0.77	0.4405
S10	[pt]: 14.66	[ps]: 11.05		0.712	0.4781
	[pt]: 14.66		**[pl]: -14.53**	**4.68**	**0.0001**
		[ps]: 11.05	**[pl]: -14.53**	**4.07**	**0.0001**

Table B.7: C2 Manner—Intergestural lag estimates (ms) of C2 manner for front-to-back sequences [pt ps pl], by speaker. **Bolded** data are significant at $p<0.05$. <u>Underlined</u> data indicate lag differences that are marginally significant.

Speaker	Plosive C2	Fricative C2	Lateral C2	*t*	*p*
	[kt]: 26.00	[ks]: 13.88		1.80	0.0742
S01	[kt]: 26.00		[kl]: 34.02	0.97	0.3340
		[ks]: 13.88	**[kl]: 34.02**	**2.42**	**0.0171**
	[kt]: 34.85	**[ks]: 19.45**		**2.15**	**0.0338**
S02	[kt]: 34.85		[kl]: 22.22	1.45	0.1501
		[ks]: 19.45	[kl]: 22.22	0.32	0.7473
	[kt]: 58.65	**[ks]: 21.33**		**3.79**	**0.0002**
S03	[kt]: 58.65		[kl]: 61.12	0.20	0.8404
		[ks]: 21.33	**[kl]: 61.12**	**3.26**	**0.0015**
	[kt]: 46.67	**[ks]: 15.55**		**4.92**	**0.0001**
S05	[kt]: 46.67		[kl]: 33.32	1.70	0.0916
		[ks]: 15.55	**[kl]: 33.32**	**2.26**	**0.0253**
	[kt]: 95.00	[ks]: 105.61		1.12	0.2649
S06	[kt]: 95.00		[kl]: 102.78	0.66	0.5094
		[ks]: 105.61	[kl]: 102.78	0.24	0.8104
	[kt]: 53.82	[ks]: 45.00		1.12	0.2629
S07	[kt]: 53.82		[kl]: 51.39	0.25	0.8022
		[ks]: 45.00	[kl]: 51.39	0.66	0.5088
	[kt]: 54.00	[ks]: 43.33		1.57	0.1196
S08	**[kt]: 54.00**		**[kl]: 29.16**	**2.94**	**0.0039**
		[ks]: 43.33	[kl]: 29.16	1.68	0.0959
	[kt]: 58.00	[ks]: 53.34		0.93	0.3556
S10	[kt]: 58.00		[kl]: 56.95	0.17	0.8665
		[ks]: 53.34	[kl]: 56.95	0.58	0.5640

Table B.8: C2 Manner—Intergestural lag estimates (ms) of C2 manner for back-to-front sequences [kt ks kl], by speaker. **Bolded** data are significant at $p<0.05$.

Speaker	Interaction		t	p
S01	**Place Order*C1 manner**	**plosive C1 + fricative C1**	**2.007**	**0.0475**
	Place Order*C2 manner	plosive C2 + fricative C2	0.846	0.3991
		plosive C2 + lateral C2	0.497	0.6204
		fricative C2 + lateral C2	0.194	0.8466
S02	Place Order*C1 manner	plosive C1 + fricative C1	1.659	0.1006
	Place Order*C2 manner	plosive C2 + fricative C2	0.428	0.6694
		plosive C2 + lateral C2	0.811	0.4192
		fricative C2 + lateral C2	0.456	0.6490
S03	Place Order*C1 manner	plosive C1 + fricative C1	0.059	0.9529
	Place Order*C2 manner	plosive C2 + fricative C2	1.884	0.0620
		plosive C2 + lateral C2	0.609	0.5439
		fricative C2 + lateral C2	0.886	0.3773
S05	**Place Order*C1 manner**	**plosive C1 + fricative C1**	**3.743**	**0.0003**
	Place Order*C2 manner	**plosive C2 + fricative C2**	**2.240**	**0.0270**
		plosive C2 + lateral C2	0.956	0.3408
		fricative C2 + lateral C2	**2.760**	**0.0067**
S06	**Place Order*C1 manner**	**plosive C1 + fricative C1**	**3.985**	**0.0001**
	Place Order*C2 manner	plosive C2 + fricative C2	1.173	0.2431
		plosive C2 + lateral C2	0.316	0.7524
		fricative C2 + lateral C2	1.262	0.2096
S07	Place Order*C1 manner	plosive C1 + fricative C1	0.532	0.5960
	Place Order*C2 manner	plosive C2 + fricative C2	0.177	0.8601
		plosive C2 + lateral C2	1.218	0.2257
		fricative C2 + lateral C2	1.365	0.1749
S08	**Place Order*C1 manner**	**plosive C1 + fricative C1**	**2.719**	**0.0078**
	Place Order*C2 manner	plosive C2 + fricative C2	0.538	0.5914
		plosive C2 + lateral C2	1.041	0.3002
		fricative C2 + lateral C2	0.614	0.5407
S10	**Place Order*C1 manner**	**plosive C1 + fricative C1**	**4.341**	**0.0001**
	Place Order*C2 manner	plosive C2 + fricative C2	0.147	0.8836
		plosive C2 + lateral C2	**3.191**	**0.0018**
		fricative C2 + lateral C2	**3.298**	**0.0013**

Table B.9: Interactions for *Place Order* and *C1 manner* and for *Place Order* and *C2 manner*, by speaker. **Bolded** data are significant at $p<0.05$.

BIBLIOGRAPHY

Baayen, H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press. Chapter 7, pp. 241–302.

Bates, D., Maechler., M., & Bolker, B. 2011. lme4: Linear mixed-effects models using S4 classes, R package version 0.999999-2. http://CRAN.R-project.org/package=lme4 (Last viewed July 3 2013).

Benus, S. & Gafos, A. 2007. Articulatory characteristics of Hungarian 'transparent' vowels. *Journal of Phonetics*, **35**, 217–300.

Browman, C. & Goldstein, L. 1992. Articulatory Phonology: An overview. *Phonetica*, **49**, 155–180.

Byrd, D. 1992. Perception of assimilation in consonant clusters: a gestural model. *Phonetica*, **49**, 1–24.

Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, **24**, 209–244.

Chitoran, I. 1999. Accounting for sonority violations: the case of Georgian consonant sequencing. *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS)*, J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A.C. Bailey (eds.), San Francisco, CA, 1999, pp. 101–104.

Chitoran, I. & Cohn, A. 2009. Complexity in phonetics and phonology: gradience, categoriality, and naturalness. In *Approaches to Phonological Complexity*, F. Pellegrino, E. Marisco, I. Chitoran, & C. Coupé (eds.), pp. 21–46. *Phonology and Phonetics*, vol. 16. Berlin, New York: de Gruyter.

Chitoran, I. & Goldstein, L. 2006. Testing the phonological status of perceptual recoverability: Articulatory evidence from Georgian. Poster presented at *Laboratory Phonology 10*, Paris, France, June–July 2006.

Chitoran, I., Goldstein, L., & Byrd, D. 2002. Gestural overlap and recoverability: articulatory evidence from Georgian. In *Laboratory Phonology 7*, C. Gussenhoven & N. Warner (eds.), pp. 419–447. Berlin: de Gruyter.

Davidson, L. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, **34**, 104–137.

Dorman, M., Studdert-Kennedy, M., & Raphael, L. 1977. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context dependent cues. *Perception and Psychophysics*, **22**, 109–122.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. 1999. Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, **25**(6), 1568–1578.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. 2001. New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, **16**(5-6), 491–505.

Fowler, C. & Galantucci, B. 2005. The relation of speech perception and speech production. In *The Handbook of Speech Perception*, D. Pisoni & R. Remez (eds.), pp. 633–652. Malden, MA: Blackwell Publishing.

Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. 2006. Towards universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, **34**(1), 49–72.

Goldstein, L., Byrd, D., & Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In *Action to Language via the Mirror Neuron System*, M. Arbib (ed.), pp. 215–249. Cambridge: Cambridge University Press.

Goldstein, L. & Fowler, C. 2003. Articulatory Phonology: A phonology for public language use. In *Phonetics and Phonology in Language Comprehension and Production*, N. Schiller & A. Meyer (eds.), pp. 159–207. Berlin: Mouton de Gruyter.

Guenther, F. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, **102**, 594–621.

Guenther, F. Ghosh, S., & Tourville, J. 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain & Language*, **96**, 280–301.

Guenther, F., Hampson, M., & Johnson, D. 1998. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, **105**, 611–633.

Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. 1998. Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, **24**(2), 592–608.

Hardcastle, W.J. & Roach, P. 1979. An instrumental investigation of coarticulation in stop consonant sequences. In *Current Issues in the Phonetic Sciences*, H. Hollien & P. Hollien (eds.), pp. 531–540. Amsterdam: John Benjamins.

Kewley-Port, D. 1983. Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **73**, 322–335.

Kochetov, A. 2006. Syllable position effects and gestural organization: articulatory evidence from Russian. In *Laboratory Phonology 8: Varieties of Phonological Competence*, L. Goldstein, D. Whalen, & C. Best (eds.), pp. 565–588. Berlin: de Gruyter.

Ktori, M., van Heuven, W.J.B., & Pitchford, N.J. 2008. GreekLex: A lexical database of Modern Greek. *Behavior Research Methods*, **40**(3), 773–783.

Kühnert, B., Hoole, P., & Mooshammer, C. 2006. Gestural overlap and C-center in selected French consonant clusters. *Proceedings of the 7th International Seminar on Speech Production (ISSP)*, H.C. Yehia, D. Demolin, & R. Laboissière (eds.), Ubatuba, Brazil, Dec. 13–15, 2006, pp. 327–334.

Liberman, A. 1954. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Journal of Experimental Psychology*, **52**, 127–137.

Lin, S. 2011. *Production and perception of prosodically varying inter-gestural timing in American English laterals*. University of Michigan, PhD dissertation. Ann Arbor, MI.

Löfqvist, A. 2010. Theories and models of speech production. In *The Handbook of Phonetic Sciences*, W. Hardcastle, J. Laver, & F. Gibbon (eds.), pp. 353–377.

Marslen-Wilson, W.D. 1987. Functional parallelism in spoken word recognition. *Cognition*, **25**, 71–102.

Mielke, J., Olson, K., Baker, A., & Archangeli, D. 2011. Articulation of the Kagayanen interdental approximant: An ultrasound study. *Journal of Phonetics*, **39**, 403–412.
Mooshammer, C., Hoole, P., & Kühnert, B. 1995. On loops. *Journal of Phonetics*, **23**(1), 3–21.

Moreton, E. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, **84**, 55–71.

Proctor, M. 2009. *Gestural Characterization of a Phonological Class: the Liquids*. Yale University, PhD dissertation. New Haven, CT.

Protopapas, A., Tzakosta, M., Chalamandaris, A., & Tsiakoulis P. 2012. IPLR: an online resource for Greek word-level and sublexical information. *Language Resources and Evaluation*, **46**(3), 449–459.

Peng, S.-h. 1996. *Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi*. Ohio State University, PhD dissertation. Columbus, OH.

Shaw, J., Gafos, A., Hoole, P., & Zeroual, C. 2009. Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology*, **26**, 187–215.

Silverman, D. 1995. *Phasing and Recoverability*. UCLA, PhD dissertation. Garland Publishing, 1997.

Sproat, R. & Fujimura O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, **21**, 291–311.

Surprenant, A. M. & Goldstein, L. M. 1998. The perception of speech gestures. *Journal of the Acoustical Society of America*, **104**, 518–529.

Wrench, A. & Scobbie, J. 2003. Categorising vocalisation of English /l/ using EPG, EMA and ultrasound. In *Proceedings of the 6th International Seminar on Speech Production (ISSP)*, S. Palethorpe & M. Tabain (eds.), Sydney, Australia, 2003, pp. 314–316.

Wright, R. 1996. *Consonant clusters and cue preservation in Tsou*. UCLA, PhD dissertation. Dissertations in Linguistics, **20**.